

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/78855>

Copyright and reuse:

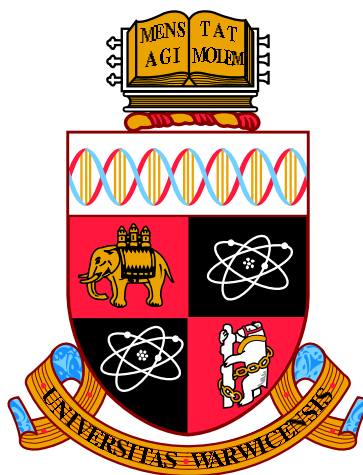
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



**Biophysical Study into the Structure and
Substrate Binding Properties
of Peptido-Mimetic Ligands to
Carboxypeptidase G2.**

by

Dhadchayini Jeyaharan

A thesis submitted in partial fulfilment for the award of
Doctor of Philosophy in Chemistry with Industrial Collaboration

Department of Chemistry

University of Warwick

September, 2015

THE UNIVERSITY OF
WARWICK

This thesis is dedicated to my parents, Thana and Jeyaharan, for their endless support and love.

Contents

Acknowledgements	ix
Declaration	x
Abstract	xi
Abbreviations	xii
List of Figures	xvi
List of Tables	xxii
1 Introduction	1
1.1 Discovery of Carboxypeptidase G2 via Anti-Folate Therapy. . . .	1
1.2 Carboxypeptidase G2 (CPG2).	3
1.2.1 Structure.	3
1.2.2 Essential Role of Zinc in Catalysis.	5
1.2.3 Potential Specific Intermolecular Interactions in MTX Hydrolysis by CPG2.	9
1.3 Biological Implications: ADEPT	10
1.3.1 From Thiocarbamate Inhibitors to the Developement of Peptido-Mimetic Inhibitors.	13

1.4	Summary.	14
1.5	Nuclear Magnetic Resonance Spectroscopy.	15
1.5.1	Quantum Angular Momentum.	16
1.5.2	Free Precession.	18
1.5.3	NMR Signal.	20
1.5.4	Nuclear Shielding.	21
1.5.5	The Notorious Size-Limitation.	24
1.5.6	Transverse Relaxation-Optimised Spectroscopy (TROSY).	25
1.6	NMR Spectroscopy in the Determination of the Binding Interface.	27
1.6.1	Saturation Transfer: a Ligand-Based Approach.	27
1.6.2	Chemical Shift Perturbation: a Protein-based Approach.	29
1.7	Research Aims.	31
2	Materials and Methods	33
2.1	Materials.	33
2.1.1	Chemicals and reagents.	33
2.1.2	Plasmid Constructs and Bacterial Strains.	34
2.1.3	Primers.	36
2.1.4	Antibodies.	38
2.2	Methods.	38
2.2.1	Cloning and DNA Conservation.	38
2.2.2	Bacterial Growth and Conservation.	44
2.2.3	Preparation of CPG2.	46
2.2.4	Preparation of Isotopically Labelled CPG2.	47
2.2.5	Protein Expression.	48

2.2.6	Protein Purification.	50
2.2.7	Quantification of Protein Concentration.	52
2.2.8	Protein Detection.	54
2.2.9	Protein Folding Assessment.	63
2.2.10	NMR Experiments.	66
2.2.11	X-ray Crystallography.	69
3	Protein Preparation for NMR Spectroscopy and X-ray Crystallography Studies	74
3.1	Introduction.	74
3.2	Expression of Codon-Optimised Full Length CPG2 in <i>E. coli</i> Leads to Intermittent Activity and Existence of Two CPG2 Subspecies. .	75
3.2.1	Testing Different <i>E. coli</i> Strains for Higher Solubility. . . .	76
3.2.2	Effect of Changes in Temperature, IPTG Concentration and Induction Duration.	77
3.2.3	Presence of Possibly Two CPG2 Isoforms.	79
3.3	Production of Soluble and Active CPG2.	81
3.3.1	Removal of the Signal Peptide Induces Expression of a Single, Soluble CPG2 Species.	81
3.3.2	Purification of CPG2 ₂₃₋₄₁₅ by Gel Filtration Chromatography.	84
3.4	High-Yield Expression of Isotopically Labelled Protein for NMR Studies.	86
3.5	Stability of CPG2 ₂₃₋₄₁₅ is Dependent Upon Buffer Conditions. . .	93
3.6	Preparation of Co-Crystals for X-ray Crystallography.	99
3.6.1	Protein Preparation.	99

3.6.2	Co-Crystallisation Attempts.	100
3.7	Conclusion.	104
4	Characterisation of Ligand Binding Properties using NMR Spectroscopy	108
4.1	Introduction.	108
4.2	CPG2 Inhibitors Used for NMR Studies.	109
4.3	Ligand-Based NMR Experiment: STD-NMR.	114
4.3.1	Use of D ₂ O in Sample Preparation is Essential.	117
4.3.2	Saturation Dependence of the STD Intensities.	118
4.3.3	Ligand Group Epitope Mapping (GEM).	121
4.3.4	Better Assessment of the Absolute Magnitude of the Saturation Effect.	127
4.4	Protein-detected NMR experiments.	137
4.4.1	¹ H- ¹⁵ N-HSQC of Free CPG2.	137
4.4.2	TROSY-NMR Characterisation of 42 kDa CPG2 ₂₃₋₄₁₅ at Two Fields and Two Temperatures.	140
4.4.3	Unravelling Binding by Using Simple HSQC NMR Experiments.	144
4.5	Conclusion	151
5	Site-directed Mutagenesis of Positions Within or Close to the Regions of CPG2 Directly Related to Substrate Binding	156
5.1	Introduction.	156
5.2	Site-Directed Mutagenesis, Expression and Purification of CPG2 Mutants.	160

5.3	Investigation of Mutant Proteins Secondary Structure Using Circular Dichroism.	166
5.3.1	Secondary Structure Differences.	166
5.3.2	Thermal Stability.	168
5.4	Use of CPG2 Spectrophotometric Assay to Study the Effects of Residue Mutation on Substrate Binding.	172
5.5	Conclusion.	175
6	Characterisation of Engineered Individual Domains of CPG2	179
6.1	Introduction.	179
6.2	Different Domains of CPG2.	180
6.3	Cloning and Expression of the Individual Domains.	181
6.3.1	Preparation of Plasmid DNA Constructs.	181
6.3.2	Effect of Changes in Temperature and IPTG Concentration on Expression.	184
6.4	Preparation of Isolated CPG2 Domains Narrows Down Requirements for Activity.	186
6.4.1	Purification of CPG2 _{CAT} , CPG2 _{CAT'} CPG2 _{DIM} by Gel Filtration Chromatography.	189
6.5	Preparation of Isolated CPG2 Domains Leads to Greatly Improved NMR Spectra.	191
6.6	Conclusion	195
7	Discussion and Perspectives	197
7.1	First Reported Method for High-Yield Expression of Active and Soluble Mature CPG2 ₂₃₋₄₁₅	198

7.2	The Aromatic Ring and Glutamate Groups are Essential Key Features in CPG2 Inhibitors.	199
7.3	Insight into CPG2 Substrate Recognition.	203
7.4	A "Divide-and-Conquer" Approach.	205
7.5	Future Directions.	207
Appendices		211
A Study of Synthetic Peptides Mimicking β3-Loop Antibody Binding Site in hCG and LH Proteins by NMR Spectroscopy		
A.1	Introduction.	212
A.1.1	Structure Dependence in 8G5 Antibody Binding of hCG β_{66-80} Epitope.	213
A.1.2	Project Aims.	216
A.2	Materials and Methods.	217
A.2.1	Sample Preparation.	217
A.2.2	TOCSY.	218
A.2.3	NOESY.	220
A.2.4	ROESY.	222
A.2.5	DQF-COSY.	223
A.3	Discussion and Results.	224
A.3.1	^1H NMR Chemical Shift Assignment of β 3-loop Peptides. .	224
A.3.2	Sequential Assignment of MOL174, MOL179 and MOL180 peptides.	228
A.3.3	Secondary Structure Prediction using Chemical Shift Index.	235

A.3.4	Secondary Structure Prediction using $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ Coupling	
	Constants.	238
A.4	Conclusion.	241
B	Chemical shift assignment of hCG peptides	245
B.1	Chemical shift assignment of MOL174 peptide.	246
B.2	Chemical shift assignment of MOL179 peptide.	247
B.3	Chemical shift assignment of MOL180 peptide.	248
	Bibliography	249

Acknowledgments

I would like to express my gratitude to my supervisor Dr Ann Dixon for welcoming me into her group, allowing me to work on this inspiring project and the incredible opportunities she provided for research, writing and travelling. Your continuous support, guidance and patience has made this thesis feasible. Thank you!

Prof Paul Davis, Dr James Schouten and Dr Joannah Towler from Mologic Ltd. (Bedford, UK) are also thanked for help, support and collaborating on this project.

Thanks to Dr Ivan Prokes for training me on the NMR spectroscopy, Phil Aston for training me on mass spectrometer instruments, and the "Sixth Floor" for their advice – especially Dr Nikola Chmel and Dr Kasra Razmkhah.

I owe my deepest gratitude to Michael Chow for valuable discussions, support in and understanding of complex problems relating to protein expression for NMR studies.

I must thank Dr Karen Ruane for her help, expert advice and valued supervision during X-ray co-crystallisation screening in Life Sciences. It was an absolute pleasure working with and learning from you.

Thanks to the members of my advisory committee, Prof Greg Challis and Dr Claudia Blindauer, for their constructive criticism and helpful suggestions for this research.

I acknowledge past and current members of the Dixon Group and the Chemical Biology Research Facility for making my PhD experience easier and more enjoyable. Special thanks goes to Leo Bowsher, Muhammad Hasan and Hasan Tanvir Imam for their friendship and constant support in the office and the laboratory!

Importantly, my beloved parents, Thana and Jeyaharan, who were particularly indispensable to me during my life, undoubting my abilities and supporting me with their love during good and bad times. This is for you.

I would also like to thank my family and friends. The Thambiayah Sisters (Daya Chitty, Nithy Periamma and Maheswary Periamma) for being so generous and supporting me during this period. My cousins, Padma and Priya Acca, for their sisterly love and boundless support. Vimala Acca and Balamurugan Anna for attending to my parents back home in France. Dharmesh and Maria for their friendship and cooking delicious food! My beautiful friends, especially Caroline, Lydie, Sophie, Hoda and Sami for their amazing support and friendship throughout the years.

Declaration

The work presented in the thesis is original, and was conducted by the author, unless stated otherwise, under the supervision of Dr. Ann Dixon. It has not previously been presented for another degree.

The work was funded by the Biotechnology and Biological Sciences Research Council (*BBSRC*) to Dr. Ann Dixon and Prof. Paul Davis.

Collaborative work has been carried out contributing towards the material presented in the thesis.

Dr. Suzanne Slade, School of Life Sciences, University of Warwick, performed tryptic digest, protein identification and analysis by Nano LC-ESI-MS/MS contributing to Chapter 2.

Philip Aston (Department of Chemistry, University of Warwick) performed analysis by Bruker MaXis II mass spectrometer contributing to Chapter 2.

Dr. Karen Ruane, School of Life Sciences, University of Warwick, tested crystals for diffraction, collected the data at Diamond Synchrotron Source in Oxford and solved the structure contributing to Chapter 3.

Dr. Dean Rea, School of Life Sciences, University of Warwick, set up sitting drop screening using Mosquito LCP crystallisation robot contributing to Chapter 3.

Dr. Sara Whittaker, School of Cancer Sciences, University of Birmingham, collected the NMR data using the 900 MHz Bruker Instrument contributing to Chapter 4.

Carla Brackstone, Department of Chemistry, University of Warwick, carried out site-directed mutagenesis of CPG2 mutants during her MChem project contributing to Chapter 5.

Dr. James Schouten and Dr. Joannah Towler, Mologic Ltd., Bedford, produced the CPG2 inhibitors and the hCG/ β 3-loop peptides (MOL174, MOL179 and MOL180) contributing to Chapters 4 and 5, and Chapter 8 respectively.

Abstract

The antibody directed enzyme pro-drug therapy (*ADEPT*) is a new anticancer treatment, where pre-clinical experiments concluded that an intermediate step involving the inhibition of carboxypeptidase G2 (*CPG2*) from the circulatory system prior to pro-drug administration is crucial to prevent systemic toxicity (Bagshawe et al. 1991). The research described in this thesis sought to better understand the mode of binding of these inhibitors to CPG2 using solution state nuclear magnetic resonance (*NMR*) spectroscopy.

A high-yield expression of active and soluble mature CPG2 (in the absence of the leader peptide) in *E. coli* suitable for NMR studies and co-crystallisation screening is reported. We have used this method to routinely produce milligram quantities of $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ isotopically-labelled protein suitable for NMR studies. The second aim of this thesis is interactions; interactions between CPG2 and selected inhibitors provided by our industrial partner Mologic Ltd. (Bedford, UK). Different structural parts of the inhibitors were identified by NMR to directly interact with CPG2: the naphthalene and the glutamate groups. Chemical shift perturbations studies show different patterns for CP06 and CP67 inhibitors suggesting that they have different binding mechanisms. Site-directed mutagenesis of residues in P1 pocket of CPG2 reveal no activity against methotrexate (*MTX*), suggesting that they are key players in substrate recognition, while H285A and E200A mutant proteins display similar activity to wild type CPG2 protein.

Although the NMR data described here for CPG2 were incomplete and thus did not yield resonance assignment, we show attempts at a "divide-and-conquer" approach. The CPG2_{CAT} construct shows great promise for downstream NMR studies as it has favourable solution properties and retains key properties of the parent protein, namely enzymatic activity and the ability to self associate.

Abbreviations

1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
AAP	<i>Aeromonas proteolytica</i> aminopeptidase
ACN	Acetonitrile
ADEPT	Antibody-Directed Enzyme Pro-drug Therapy
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
Amp	Ampicillin
APS	Ammonium Persulfate
Cam	Chloramphenicol
CD	Circular Dichroism
COSY	COrrrelation SpectroscopY
CPG2	Carboxypeptidase G2
CSA	Chemical Shift Anisotropy
DD	Dipole-Dipole
DHFR	Dihydrofolate Reductase
DNA	Deoxyribo-Nucleic Acid
dNTPs	Deoxynucleotide Triphosphates

DQF-COSY	Double-quantum filtered COSY
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
FA	Formic Acid
Fab	Antigen Binding Fragment
FID	Free Induction Decay
Fmoc	9-Fluorenylmethyloxycarbonyl
FSH	Follicle-Stimulating Hormone
hCG	Human Chorionic Gonadotropin
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear Single Quantum Correlation
IgG	Immunoglobulin G
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
3J	three-bond J -coupling
Kan	Kanamycin
LB	Lysogeny Broth
LC-ESI UHR QTOF	Liquid Chromatography Electrospray Ionisation Ultra-High Resolution Quadrupole Time-Of-Flight
LC-ESI-MS/MS	Liquid Chromatography Electrospray Ionisation Tandem Mass Spectrometry
LH	Luteinizing Hormone

MALDI-TOF MS	Matrix Assisted Laser Desorption / Ionisation Time Of Flight Mass Spectrometry
MOPS	3-(N-morpholino)propanesulfonic acid
MWCO	Molecular Weight Cut-Off
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OD	Optical Density
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PIPES	piperazine-N,N-bis(2-ethanesulfonic acid)
RF	Radio Frequency
ROE	NOE in a rotating frame
rpm	revolutions per minute
SAM	S-Adenosylmethionine
SDS-PAGE	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
SPR	Surface Plasmon Resonance
SPR	revolutions per minute
STD	Saturation Transfer Difference
STD-AF	STD Amplification Factor
T ₁	Spin-lattice (or longitudinal) relaxation time
T ₂	Spin-spin (or transverse) relaxation time
TAE	Tris-Acetate-EDTA
TBST	Tris-Buffered Saline and Tween-20

TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TOCSY	Total Correlation Spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
TROSY	Transverse Relaxation Optimised Spectroscopy
TSH	Thyroid-Stimulating Hormone
WaterLOGSY	Water Ligand Observation with Gradient Spectroscopy
2×YT	2×Yeast extract and Tryptone medium

List of Figures

1.1	MTX hydrolysis by CPG2.	3
1.2	Ribbon diagram of CPG2 tetramer and dimer.	4
1.3	X-ray crystallography structure of CPG2.	6
1.4	Conserved metal-binding domains of CPG2, AMP and GCPII. . .	8
1.5	Key amino acids in CPG2 in MTX recognition.	10
1.6	Basic principle of ADEPT.	11
1.7	CPG2 inhibitors.	14
1.8	Spin polarisation.	17
1.9	The Zeeman effect.	18
1.10	Free precession.	19
1.11	Radio frequency pulse.	20
1.12	NMR signal.	21
1.13	Nuclear shielding.	23
1.14	NOE effect.	28
1.15	Saturation transfer difference NMR spectroscopy.	29
1.16	Exchange regime.	31
2.1	Elution and Calibration curve obtained with β -amylase, Albumin and Cytochrome C protein Standards run on HiLoad 16/600 Superdex 75 g GL column.	59

2.2	Hydrolysis of MTX measured at 320 nm.	65
2.3	Different vapour diffusion system.	71
3.1	Expression in different <i>E. coli</i> strains.	77
3.2	Testing different IPTG induction concentration and expression duration at 15°C.	78
3.3	Expression of CPG2 in the pet28a vector revealing the presence of two immature isoforms.	80
3.4	Agarose gel electrophoresis (2%) of PCR products.	82
3.5	Expression and characterisation of CPG2 ₂₃₋₄₁₅	84
3.6	Calibration curve obtained with Gel Filtration standards on HiLoad 16/600 Superdex 75 g GL column.	86
3.7	Comparison of CPG2 expression in 2×YT and M9 media.	87
3.8	Growth curves of cultures post-inoculation with minimal (M9) and rich 2×YT media at different temperatures.	88
3.9	SDS-PAGE gels of double colony selection and condensation method.	90
3.10	Maldi-TOF mass spectrometry <i>A</i>) labelled and <i>B</i>) unlabelled CPG2 ₂₃₋₄₁₅	92
3.11	Characterisation of CPG2 ₂₃₋₄₁₅ in buffers 1-3,primarily differing in their NaCl content.	95
3.12	Characterisation of CPG2 ₂₃₋₄₁₅ in 50 mM Arg / 50 mM Glu.	96
3.13	Crystal grown using conditions described in table.	102
3.14	Thin needle of protein crystals.	103
4.1	CP06 and CP67 inhibitors produced by Mologic Ltd.	112
4.2	CP39 and CP42 inhibitors produced by Mologic.	113

4.3	STD-NMR: 3 step concept.	115
4.4	Testing on-resonance frequencies.	116
4.5	Buffer conditions for STD-NMR experiments.	118
4.6	Saturaton time t_{sat} build-up curves.	120
4.7	STD NMR spectrum of 1.59 mM CP06 in the presence of 20 μM CPG2 recorded at 298 K on a 700 MHz spectrometer.	122
4.8	STD NMR spectrum of 1.59 mM CP67 in the presence of 20 μM CPG2 recorded at 298 K on a 700 MHz spectrometer.	124
4.9	STD NMR spectrum of 1.59 mM CP42 in the presence of 20 μM CPG2 recorded at 298 K on a 700 MHz spectrometer.	126
4.10	STD-NMR titration experiments of CPG2 by inhibitors.	128
4.11	Binding isotherms fits for CP06.	131
4.12	Binding isotherms fits for CP67.	133
4.13	Binding isotherms fits for CP42.	134
4.14	Three CPG2 ^1H - ^{15}N -HSQCs at different temperatures and number of scans.	139
4.15	^1H - ^{15}N TROSY-HSQC spectra acquired at different fields and temperature.	142
4.16	^1H - ^{15}N TROSY-HSQC spectrum of $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ -labelled CPG2 ₂₃₋₄₁₅	143
4.17	^1H - ^{15}N -HSQC spectra of CPG2 acquired in the presence of CP06 ligand.	145
4.18	Close-up of ^1H - ^{15}N -HSQC spectra of CPG2 acquired in the presence of CP06 ligand.	146
4.19	CP06 curve fits.	148

4.20	^1H - ^{15}N HSQC spectra of CPG2 acquired in the presence of CP67 ligand.	150
4.21	Close-up of ^1H - ^{15}N HSQC spectra of CPG2 acquired in the presence of CP67 ligand.	151
5.1	Zinc binding domain.	157
5.2	Two possible pockets in the CPG2 active site.	159
5.3	Residues mutated to alanine in the active site.	161
5.4	Typical purification of a mutant construct.	163
5.5	Purification of mutants.	164
5.6	Coomassie-stained SDS-PAGE gel comparison of expression levels of wild type and point mutant proteins.	165
5.7	Circular dichroism.	167
5.8	Melting temperature determined by CD melts at 220 nm.	169
5.9	Enzymatic activity assays of each mutant and wild-type proteins.	173
5.10	Competitive enzymatic activity assays of H385A and E200A mutant proteins.	174
6.1	Different domains of CPG2.	181
6.2	Construction of His-tagged recombinant proteins.	182
6.3	1-2% agarose gel electrophoresis of PCR products.	183
6.4	Effects of IPTG concentration and temperature changes on expression.	185
6.5	Characterisation of CPG2 _{CAT} , CPG2 _{CAT'} and CPG2 _{DIM}	186
6.6	MALDI-TOF spectra of the isotopically labelled proteins.	188

6.7	Calibration curve obtained with Gel Filtration standards on HiLoad 16/600 Superdex 75 g GL column.	190
6.8	^1H - ^{15}N HSQC spectra of the different domains.	192
6.9	^1H - ^{15}N HSQC spectra comparison of CPG2 _{CAT} and CPG2 ₂₃₋₄₁₅ . . .	193
6.10	CD spectra of CPG2 _{CAT}	194
7.1	Summary of the mutations and their respective activities mapped onto the X-ray crystal structure of CPG2.	205
A.1	Crystal structure of hCG.	214
A.2	Binding results from Pepscan - Mab 8G5.	215
A.3	Information provided by TOCSY experiment.	219
A.4	General pulse sequence for TOCSY.	220
A.5	Information provided by NOESY experiment.	221
A.6	General NOESY pulse sequence.	221
A.7	General ROESY pulse sequence.	222
A.8	Information provided by DQF-COSY experiment.	224
A.9	Different regions of the 2D TOCSY spectrum.	226
A.10	700 MHz 2D TOCSY spectrum of MOL174.	227
A.11	Time dependance of NOE.	229
A.12	2D ^1H NOESY spectra of MOL174 at different mixing times. . . .	230
A.13	NOE as a function of molecular tumbling rate.	232
A.14	Variation of the NOE sign in the NOESY and ROESY experiments, and ROE artifacts	233
A.15	Overlay of 2D ^1H ROESY and 2D ^1H TOCSY spectra in the amide region.	234

A.16 CSI values of H_α	237
A.17 H_N - H_α cross-peaks and cross sections parallel to F2 frequency axis from the DQF-COSY spectrum of MOL174.	238
A.18 Data summary.	243

List of Tables

2.1	Plasmid constructs used in this study.	34
2.2	<i>E. coli</i> strains used in this study.	35
2.3	Primers used in this study.	37
2.4	Cycling conditions used for PCR amplification.	39
2.5	Cloning reaction.	41
2.6	PCR screening of bacterial colonies.	43
2.7	Recommended antibiotic concentrations for rich and minimal media.	45
2.8	SDS-PAGE gels sample buffer and running buffer.	55
2.9	SDS-PAGE gel recipe.	56
2.10	Mass average of all expressed constructs (without TEV protease cleavage) obtained from LC-ESI UHR QTOF mass spectrometry.	62
3.1	Difference in expression yield between wild-type CPG2 and CPG2 ₂₃₋₄₁₅	82
3.2	Buffer composition for buffers 1-3.	93
3.3	CPG2 precipitation matrix to determine protein solubility and stability conditions for NMR studies using the X-ray screening method.	98
3.4	First attempt at co-crystallisation: screening using the hanging drop method.	101

3.5	Second trial at co-crystallisation: screening using the hanging drop method.	103
4.1	Apparent binding constants.	136
6.1	Difference in expression yield between CPG2 ₂₃₋₄₁₅ and its individually expressed domains.	187
A.1	Sequence of peptides derived from hCG β 3-loop.	216
A.2	³ J-couplings between N and C $_{\alpha}$ protons, and their respective Φ dihedral angle in all three peptides.	240

Chapter 1

Introduction

1.1 Discovery of Carboxypeptidase G2 via Anti-Folate Therapy.

Reduced folates play a pivotal role as co-enzymes in many biochemical pathways. DNA synthesis is one of the most important pathways reduced folates are implicated in, and this metabolic pathway can be regulated through two different routes. The first is via DNA precursors, thymidine and purines, whose synthesis are dependant on folate co-enzymes (Bertino and Hillcoat 1968). The second is by way of DNA methylation, whereby folate co-enzymes are required in the biosynthesis of methionine (Crider et al. 2012). Methionine is needed for the synthesis of S-adenosylmethionine (*SAM*). *SAM* is a methyl group (one-carbon unit) donor operating in most biological methylation reactions, such as specific sites within the DNA sequence. The DNA

methylation takes part in controlling gene expression, and anomalies in DNA methylation have been associated with the development of cancer. As a result, folic acid is a target molecule for chemotherapy and its depletion has been accomplished by the use of folic acid antagonists, specially 4-amino-N¹⁰-methyl-pteroyl-L-glutamate – also known as methotrexate (*MTX*). This compound inhibits the dihydrofolate reductase (*DHFR*) enzyme that is essential in the biosynthesis of all folate co-enzymes (Osborn and Huennekens 1958). As a consequence, *MTX* has become one of the most widely used chemotherapeutic agents today. High-doses of methotrexate (more than 1 mg.m⁻² of total body surface area) are also used to efficiently treat human malignancies and diseases, such as leukaemia and rheumatoid arthritis. Although high-dose methotrexate carries a low risk for severe treatment-related toxicity, 2-10% of the patients experience methotrexate-related toxicity that could hold up further treatment or may be life threatening (Schwartz et al. 2007). Hence work had started on enhancing methotrexate clearance in rescue therapies.

Depletion of *MTX* was achieved directly by enzymatic cleavage of the *MTX* peptide bond to remove the glutamate moiety. The enzymes capable of such hydrolysis were named carboxypeptidase G, and various forms have been isolated throughout the years. Actually, shortly following the exploitation of methotrexate in anti-cancer therapy, a loss of biological activity (i.e. *DHFR* inhibition by *MTX*) was detected after contact with bacteria. In 1967, a bacterial enzyme, which catalyses the cleavage of a glutamate moiety from folic acid derivatives, was isolated (carboxypeptidase G). Figure 1.1 shows the

products produced from hydrolysis of MTX, 2,4-diamino-N₁₀-methylpteroic acid (*DAMPA*) and glutamate, which were less active inhibitors of DHFR (Valerino et al. 1972). Carboxypeptidase G1 (*CPG1*), extracted from a strain of *Pseudomonas stutzeri*, was used to lower methotrexate blood concentration because of its 100-fold greater affinity for methotrexate than carboxypeptidase G. However, its bacterial source was misplaced (Chabner et al. 1972). Nonetheless, Minton et al. (1983) managed to isolate carboxypeptidase G2 (*CPG2*) from *Pseudomonas sp. Strain RS-16* by cloning the gene into a pBR322 vector and transforming it in *Escherichia coli*. The expression method allowed sufficient production of CPG2 to prevent methotrexate toxicity at a therapeutic level (Sherwood et al. 1985). Compared to its predecessors CPG and CPG1, CPG2 has a much higher affinity for methotrexate.

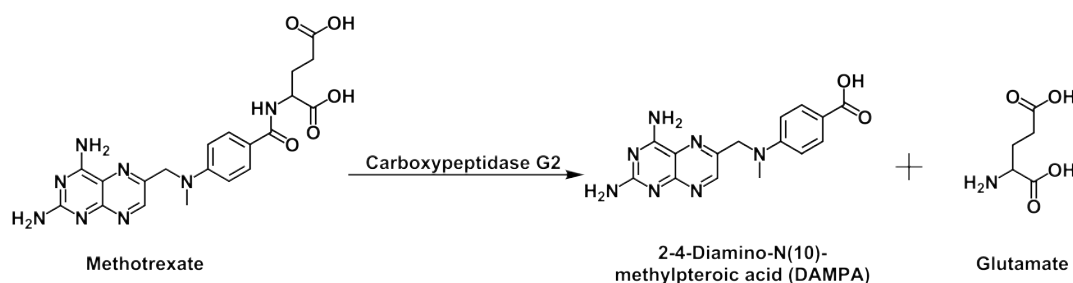


Figure 1.1: MTX hydrolysis by CPG2. Cleavage of methotrexate results in the formation of DAMPA and glutamate: two non-toxic compounds.

1.2 Carboxypeptidase G2 (CPG2).

1.2.1 Structure.

CPG2 is a 415 amino acid zinc-dependant metalloenzyme produced as an asymmetric homodimer of molecular weight $2 \times 43,932$ Da according to the

X-ray crystal structure (Rowse et al. 1997) (Figure 1.2-A,B). Minton et al. (1984) revealed that the enzyme might have two possible signal peptidase cleavage sites producing (i) a normal CPG2 protein with an N-terminal Ala corresponding to position 23 and (ii) a truncated version of CPG2 with an N-terminal Thr corresponding to position 37 of the primary amino acid sequence (Lindner et al. 2003). Sequence analysis demonstrated that the signal peptide preceding the N-terminal alanine, targets the CPG2 to the periplasmic space of *Pseudomonas* (Sherwood et al. 1985).

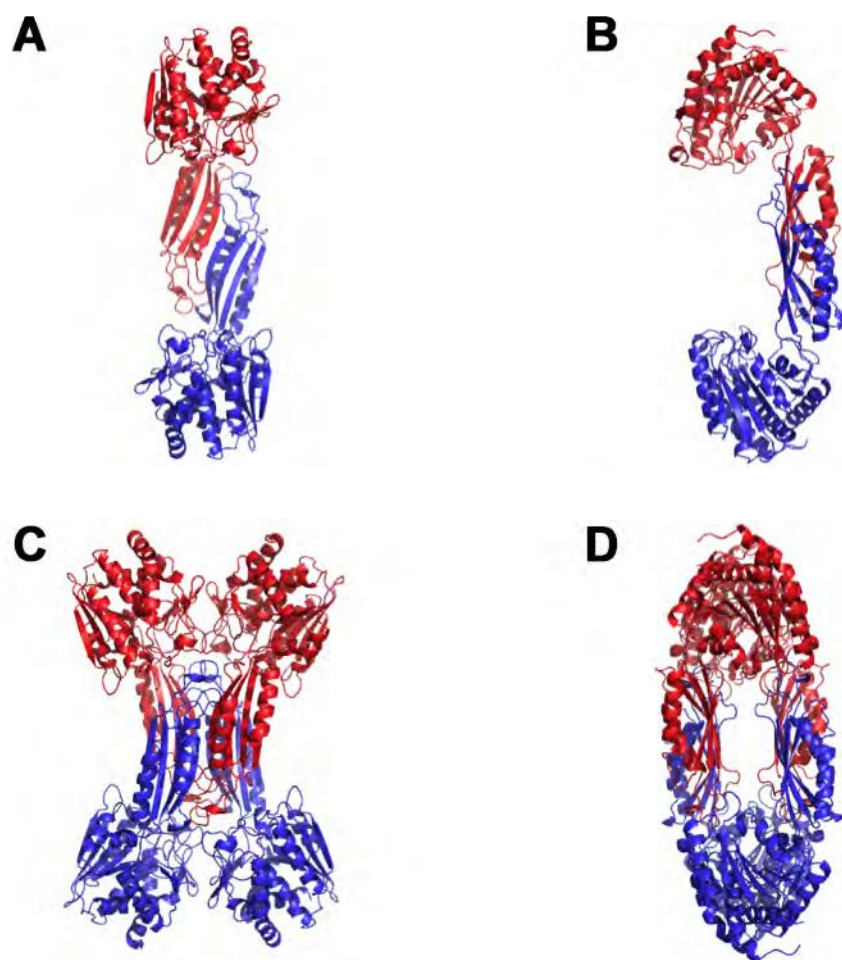


Figure 1.2: Ribbon diagram of CPG2 tetramer and dimer. One CPG2 monomer is coloured blue, while the other one is coloured in red. A) Front view of the dimer, B) Side-view of the dimer, C) Front view of the tetramer, D) Side-view of the tetramer. The CPG2 tetramer shows interactions between the β -strands of CPG2 monomers' dimerisation domains and helices of CPG2 monomers' catalytic domains, while CPG2 dimer shows the dimer interface. PDB: 1CG2, Rowsell et al. (1997).[Pymol Viewer]

The overall structure of the CPG2 monomer consists of two domains: the metal-binding domain, which likely delivers CPG2 substrates to the two zinc ions in the active site, and the dimerisation domain (Figure 1.3). The dimerisation domain, which is inserted in the middle of the metal-binding site sequence, interacts through hydrogen bonding between two β -strands, one from each monomer, forming a continuous β -sheet across the whole dimer, as well as hydrophobic interactions between helices (Rowse et al. 1997) (Figure 1.2). It is not clear whether dimerisation domain of CPG2 is necessary for catalysis to take place. The related homodimeric enzyme PepV from the M20 exopeptidase family, of which CPG2 is a member, has been shown to be enzymatically active as a monomer (Jozic et al. 2002). Additionally, the X-ray crystal structure also showed the presence of tetrameric structures due to interactions between the catalytic domains of each dimeric proteins (Rowse et al. 1997) (Figure 1.2-C,D). The question is, then, how biologically relevant these dimeric and tetrameric structures are; if they are, these oligomeric states would also persist in solution.

1.2.2 Essential Role of Zinc in Catalysis.

Depending on the number of metal ions, metalloproteins can be divided into two groups: mononuclear and co-catalytic (i.e. binuclear) enzymes. CPG2 contains two-metal ions bound in its active site, thus it belongs to the binuclear group (Figure 1.4-A). Zinc 1 is bound to Glu176 and His385, while zinc 2 is bound to Glu200 and His112. The two zinc ions are bridged by Asp141, and are

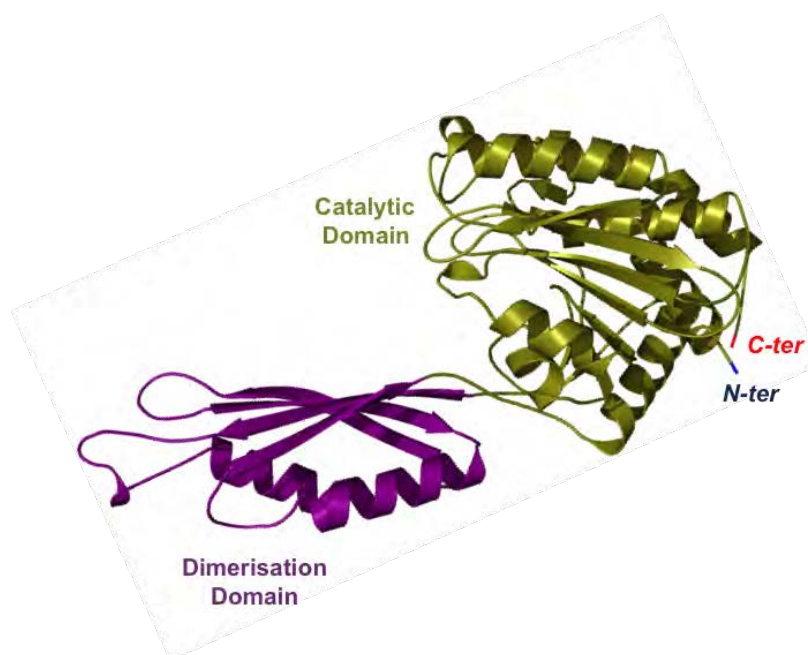


Figure 1.3: X-ray crystallography structure of CPG2. CPG2 monomer consists of two domains (i) a metal-binding domain, the active site (represented in green) and (ii) a smaller dimerisation domain (represented in purple). PDB: 1CG2, Rowsell et al. (1997). [Pymol Viewer]

held in a tetrahedral coordination in complex with histidine, glutamic acid and aspartic acid residues. Although the exact role of these metal ions (called the catalytic and co-catalytic metal ions respectively) are not fully understood yet, they display differences in terms of binding constants, spectroscopic features and evolutionary origin (Bzymek and Holz 2004; Lowther and Matthews 2002; Wouters and Husain 2001). Near the active site Glu175 forms a hydrogen bond to the bridging water molecule, and is thought to function as an acid/base during catalysis by activating the water molecule, which may be important during catalytic turnover (Figure 5.1). Analysis of the sequences of the M20 family revealed that enzymes exhibiting carboxypeptidase specificity have all a glutamate residue (residue Glu175 for CPG2) at the same position, while proteins with aminopeptidase specificity displayed an aspartic acid residue (Biagini and Puigserver 2001); it suggested that Glu175 may be important in substrate recognition too.

The members of the M20 family are known to be aminoacylases or exopeptidases, such as CPG2 (Levy and Goldman 1967). Many of these enzymes have shown potential as therapeutic agents. Succinyldiaminopimelate desuccinylase is considered as an excellent target in anti-bacterial therapy (Bienvenue et al. 2003), while CPG2 is considered as a very promising therapeutic agent in anti-cancer treatment (see Section 1.3). Also, CPG2 has very similar structure to aminopeptidase (*AMP*) and membrane-bound human glutamate carboxypeptidase II (*GCPII*). Their mechanisms of catalysis were proposed via the use of X-ray crystallography, site-directed mutagenesis and computational methods (Klusak et al. 2009, Schurer et al. 2004). The CPG2 metal-binding domain contains structural similarities to the single-domain aminopeptidase of AAP and GCPII, which is also involved in catalysis. These proteins contain all eight β -sheets and seven helices, and the active sites of all proteins superpose incredibly well (Rowell et al. 1997). Superposition of the metal-binding sites in Figure 1.4 shows that four out of the five zinc ligands are conserved in CPG2, compared to AMP and GCPII (Glu200 in CPG2 is changed for an aspartate residue in AMP and GCPII). The accepted consensus for this class of enzymes is that the bridging oxygen is provided by a hydroxide anion rather than a water molecule (crystallography showed the presence of a water molecule), and the HO^- ligand must turn from the bridging position to coordinate to Zinc 1 only before it can perform a nucleophilic attack on the carbonyl carbon of the substrate and form an anionic tetrahedral intermediate (Klusák et al. 2009), and cleavage of the carbon-nitrogen bond and product formation.

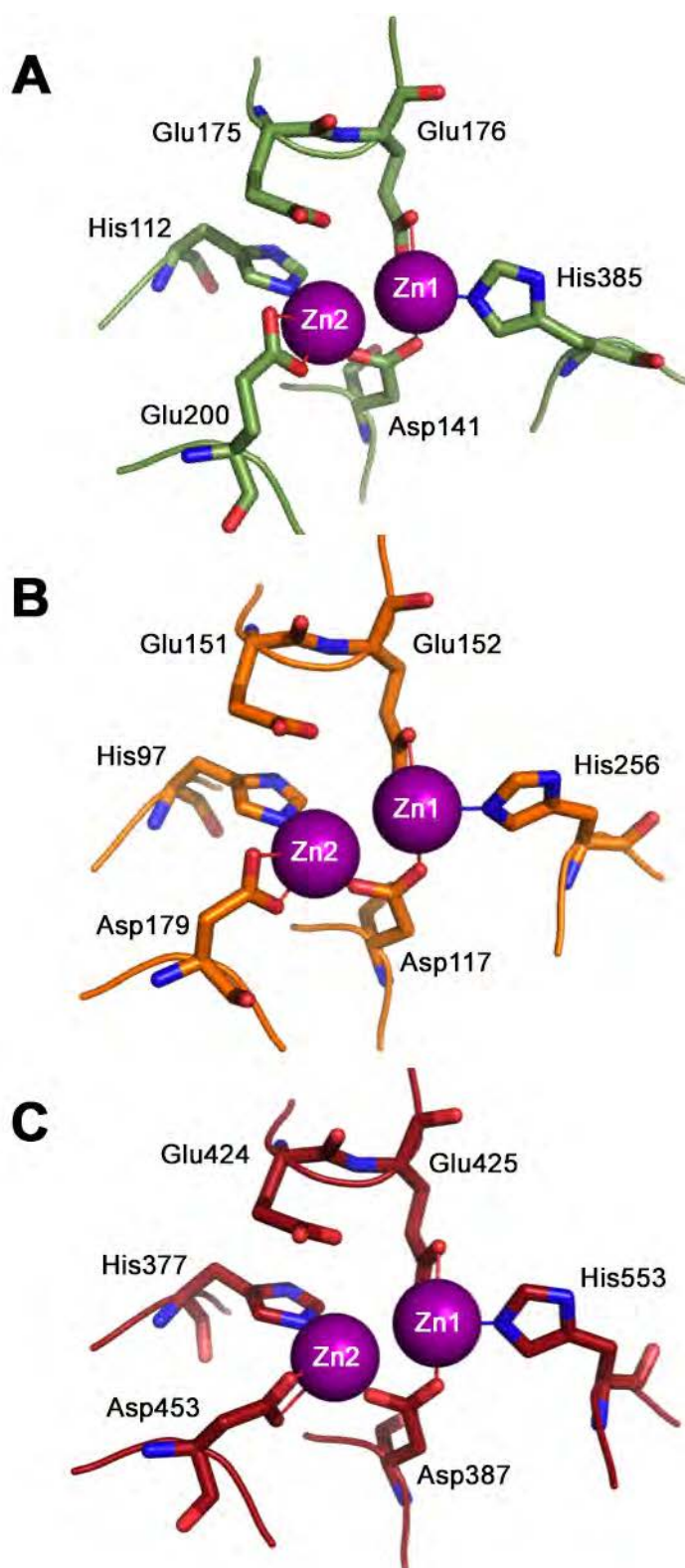


Figure 1.4: Conserved metal-binding domains of CPG2, AMP and GCPII. Active sites of *A*) CPG2, *B*) AMP and *C*) GCPII in similar orientations to show the conservations of the zinc ligands. Also shown is the glutamate residue (Glu175 in CPG2, Glu151 in AMP and Glu424 in GCPII) which is thought to facilitate the nucleophilic attack of the water on the substrate. Adapted from Rowsell et al. (1997) (PDB: 1CG2), Chevrier et al. (1994) (PDB: 1AMP) and Barinka et al. (2004) (PDB: 2OOT). [Pymol Viewer]

1.2.3 Potential Specific Intermolecular Interactions in MTX Hydrolysis by CPG2.

Using information extracted from the studies described above for related enzymes, a binding mode hypothesis was established for the CPG2-MTX complex by applying molecular docking and molecular dynamics simulation, as the mechanism of action of CPG2 has not yet been experimentally determined (Turra et al. 2012). Turra et al. (2012) suggested that MTX was interacting at very specific positions in the CPG2 catalytic region, which may be at the helm of substrate recognition and cleavage mechanism (see Figure 5.3 in Chapter 5). They described the presence of a hydrophilic pocket (P1) near the CPG2 active site accommodating the large pterate moiety of MTX required for enzyme-substrate interaction (Sherwood et al. 1985) (Figure 1.5). The pteridine moiety of MTX was found to be held via hydrogen bonding interaction between endo- and exocyclic nitrogen atoms of the pterate and the carbonyl oxygen of Lys208 and Ser210. On the other hand, the glutamate group was found to point towards the surface (P2) through the formation of a hydrogen bond with Arg324. Rowsell et al. (1997) showed that mutation of Arg324 to alanine resulted in low activity of the CPG2 mutant against MTX, but its exact role remains unexplained. To this date, no further mutation studies of CPG2 were performed. Additionally, Turra et al. (2012) also revealed the possibility of an intermolecular interaction between the amide of MTX and the carbonyl oxygen of Gly360.

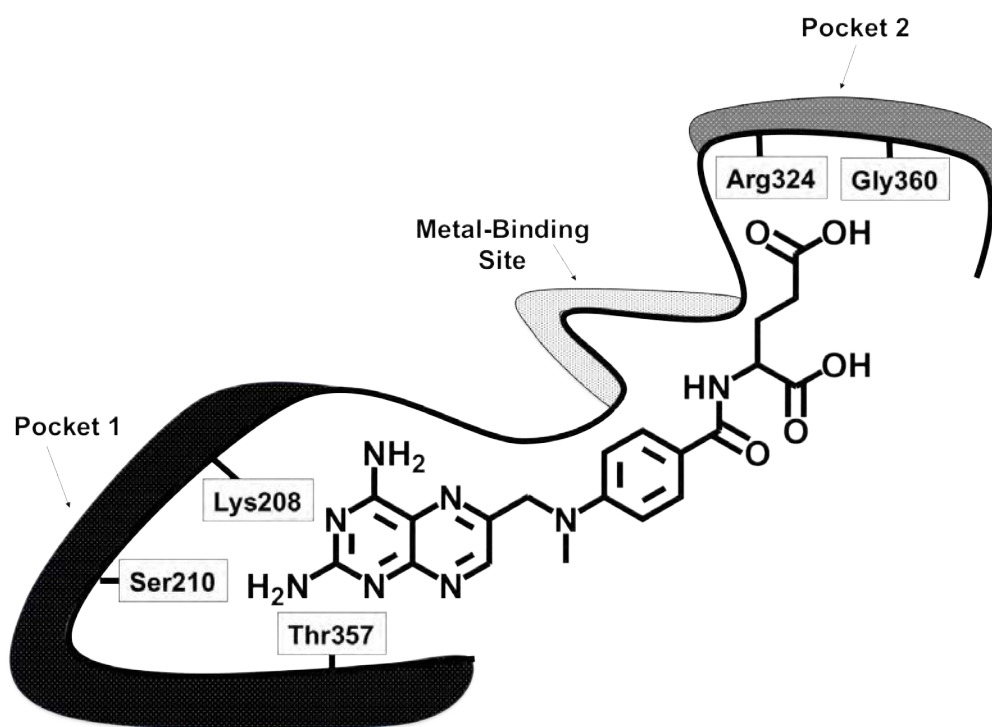


Figure 1.5: Key amino acids in CPG2 in MTX recognition. A schematic view of the catalytic site is represented, with the different amino acids, in the different pockets, thought to directly interact with substrate MTX.

1.3 Biological Implications: ADEPT

CPG2 is currently being exploited in a rescue therapy following high-dose methotrexate treatment because of its ability to metabolise the highly cytotoxic methotrexate to two non-toxic metabolites. In this project, we were mainly interested in the use of CPG2 in a new anticancer treatment: the antibody directed enzyme pro-drug therapy (*ADEPT*). It is well-known that the efficiency of chemotherapy in cancer is restricted because of its actions on healthy tissues. This complication can be prevented by restricting the effects of the cytotoxic drugs to cancer sites. However, this requires that the cytotoxic drug to be exclusively produced at tumour sites; and second, however no less essential, the cytotoxic drug should be retained at tumour sites or lose its cytotoxic ability on leaving the tumour location. Accordingly the concept of

ADEPT was put forward in the mid-1980s as a method of limiting the action of cytotoxic drugs to tumour sites, thus amplifying their effectiveness and potentially diminishing their unacceptable lethal effects on normal tissues (Bagshawe 1987). It was proposed that this could be accomplished in several steps. In the first step, CPG2 is administered as an immunoconjugate attached to an antibody directed at tumour-associated antigen (Figure 1.6-Stage 1). Once it has reached cancer sites, it can enzymatically activate a pro-drug (e.g. glutamated nitrogen mustard drug) that carries an inactivating component the glutamate, which is cleavable by CPG2 selectively at cancer sites (Figure 1.6-Stage 3). The benefit of this technique is the ability to deliver a much higher concentration of active drug to treat cancer due to the higher turnover rate of the pro-drug by the enzyme. Moreover, it would surmount the problem of heterogeneity of tumour marker expression within the tumour tissue so that neighbouring tumour cells that do not express the specific antigen would also be eradicated through the supposed bystander effect.

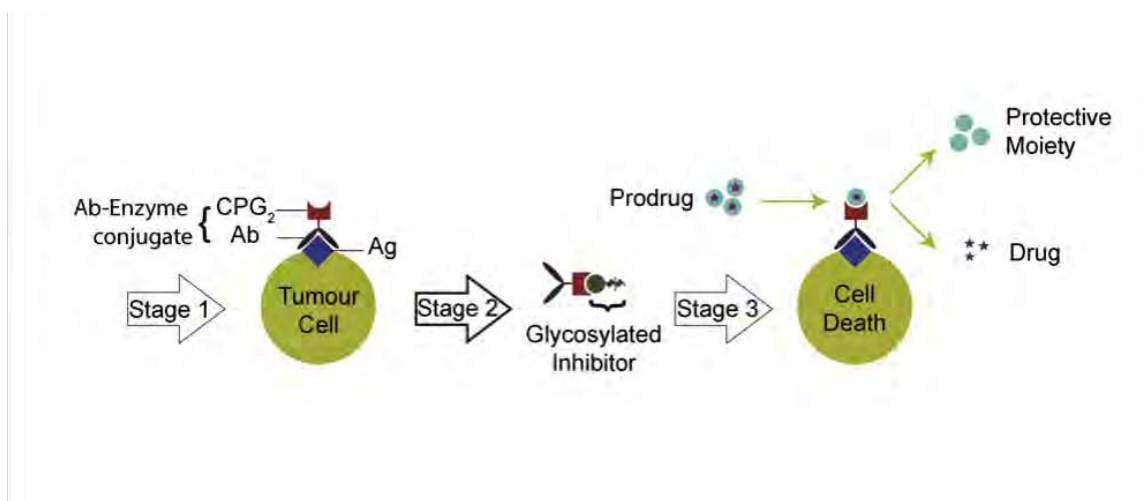


Figure 1.6: Basic principle of ADEPT. ADEPT is a 3 steps concept involving (i) stage 1: immunoconjugate administration; (ii) stage 2: inhibition and clearance of the enzyme at non-tumour sites; (iii) stage 3: pro-drug administration. Adapted from Bagshawe et al. (1990).

The choice of the enzyme raised questions of great importance. It was vital to avoid a human enzyme to prevent the pro-drug from being activated wherever it was spontaneously located. Bacterial enzymes are known to have the advantage of improved kinetics with substrate turnover, and a number of them perform most favourably under physiological conditions. As a consequence, CPG2 is ideal for use in ADEPT because it has no mammalian homologue, thus no endogenous enzymes would act on a pro-drug specific for CPG2, and being a bacterial enzyme, has the advantage of enhanced kinetics with substrate turnover (Searle et al. 1986). Although non-mammalian enzymes with no human equivalent have the benefit of singularity, immune response to CPG2 occurred in patients (Sharma et al. 1992). Understandingly the principle epitopes involved in antibody interactions were modified to diminish their immunogenicity (Spencer et al. 2002).

Despite all these improvements to reinforce the prospects of ADEPT becoming a powerful cancer treatment, pre-clinical experiments concluded that an intermediate step involving the inhibition of CPG2 from the circulatory system prior to pro-drug administration was indispensable to prevent systemic toxicity (Bagshawe et al. 1991). Thus ADEPT requires several specific constituents: 1) the antibody-enzyme conjugate, 2) the prodrug, and 3) the clearance of the enzyme (Figure 1.6). It is evident that the interactions and combination of these various elements need to be investigated. However, it is necessary to consider them individually first. The goal of this thesis is to study and characterise CPG2-inhibitors interactions important for enzyme clearance (Figure 1.6-Stage 2).

1.3.1 From Thiocarbamate Inhibitors to the Developement of Peptido-Mimetic Inhibitors.

Early inhibitors were designed to contain the structural features of CPG2 substrates (i.e. MTX). The features that seemed to be crucial include a α -carboxylated moiety on the L-glutamate residue and a benzene ring close to the carbonyl carbon of the amide bond as in MTX (Khan et al. 1999; DeAngelis et al. 1996). Khan et al. (1999) generated thiolate-containing inhibitors based on a thiocarbamate moiety attached to a benzene ring have also been reported for CPG2 but were only moderately effective competitive inhibitors and displayed toxicity towards human colonic epithelial cancer cell lines (LS174T cells), possibly allowing their use as inhibitors in ADEPT (Khan et al. 1999) (Figure 1.7-A). Hence, work had started by Mologic (Bedford, UK) to design and synthesise the second generation of inhibitors (Figure 1.7-B). They have designed peptido-mimetic ligands that have a "warhead" that binds and inhibits the enzyme, and a galactosylated tail that binds to hepatocyte receptors (i.e. asialoglycoprotein receptors) to provide clearance of CPG2 at non-tumorous locations in a patient. Typically the peptides were developed to bind and inhibit CPG2 without any significant enzyme modification (i.e. hydrolysis), and facilitate clearance of the CPG2-inhibitor complex via the liver by binding to hepatic cells expressing asialoglycoprotein cell-surface receptors, which are proteins with sialic acid residues removed from glycoproteins and receptors ensuring rapid clearance of the molecules with subterminal galactose residues.

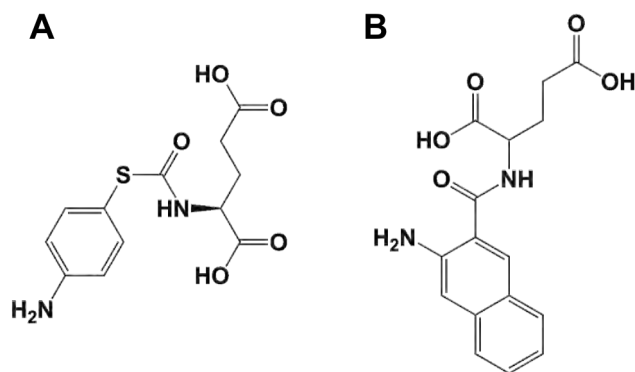


Figure 1.7: CPG2 inhibitors. A) Thiocarbamate inhibitor synthesised by (Khan et al. 1999), B) New peptido-mimetic inhibitor CP06 produced by Mologic Ltd. (Bedfore, UK).

1.4 Summary.

Due to its therapeutic applications in the treatment of cancer and autoimmune diseases, CPG2 is of great biomedical interest. The crystallographic data helped to better inform early rational drug design efforts (Rowse et al. 1997; Lindner et al. 2003), but further attempts at co-crystallisation of CPG2 with ligands, substrates or inhibitors have not been reported and very little progress has been made since these early studies to establish key features of CPG2 such as the location of its active site, the presence of additional ligand-binding sites, its stability, its oligomeric state, and the molecular basis of activity. Thus far, our current working understanding of the molecular basis of CPG2 activity is largely based on sequence / structural homology and molecular modelling (Lindner et al. 2003). Solution state nuclear magnetic resonance (*NMR*) spectroscopy is a unique tool for studying protein structure, stability, dynamics, and ligand binding properties that could inform structure-based drug discovery and provide a fundamental understanding of CPG2 function (Wuthrich 1986; Clarkson and Campbell 2003; Stark and Powers 2012).

1.5 Nuclear Magnetic Resonance Spectroscopy.

There are several methods by which one can get structural information about enzymes. For instance, gel filtration and western-blotting can tell us about the macroscopic level information of the enzyme. Site-directed mutagenesis can also be used (see Chapter 5) to study and monitor what kind of changes occurred in order to draw a conclusion about the structure via the investigation of substrate-specificity. These studies could lead us closer to the structure, but we still cannot be sure about the exact three dimensional structure. X-ray crystallography and NMR spectroscopy are the only established approaches available to resolve high resolution molecular structures. Although both methods may complement each other to get a better picture of a given protein structure, NMR is particularly suited to study of a soluble enzyme and investigation of enzyme-ligand interactions that are closer to a real functional environment (i.e. physiological conditions in solution).

The following section provides a very brief description of the NMR basics. Since NMR spectroscopy is based mostly on the quantum mechanics of nuclear spin angular momentum, which cannot be visualised by a classical route, for clarity the classical representation is going to be adopted to describe the NMR experiments.

1.5.1 Quantum Angular Momentum.

Spectroscopy is a technique that relies on the interaction between matter and electromagnetic radiation. Electrons possess an intrinsic electron spin angular momentum known as "spin". Protons and neutrons (and therefore some nuclei) also possess a spin angular momentum called the nuclear spin angular momentum I . The spin angular momentum is quantised. If a particle can spin, it has an angular momentum and therefore has quantised energy levels (1.1).

$$I = \hbar\sqrt{I(I+1)} \quad (1.1)$$

where \hbar =Plank's constant/ 2π

A circulating particle with both charge and spin angular momentum creates a current and should have an associated magnetic field. The magnetic field produced by a particle is called its magnetic moment (μ) and is related to the spin angular momentum via Equation (1.2). The proportionality constant is the gyromagnetic ratio (γ). A positive value of γ means the magnetic moment is parallel to the spin angular momentum. The value of γ is characteristic of each nuclei and result in different intensity of the magnetic moment and differences in their NMR frequencies.

$$\mu = \gamma I \quad (1.2)$$

In biomolecular NMR spectroscopy, the mainstay nuclei are ^1H , ^{13}C and ^{15}N . Hence, the NMR behaviour is based on the interactions between nuclei with a nuclear spin quantum number $I \geq 1/2$. Quantum mechanics tells us that a

nucleus with a nuclear spin quantum number I will split into $2I+1$ possible energy states. In the absence of a magnetic field, the distribution of magnetic moments of an ensemble of nuclei is isotropic (all possible orientations represented). The application of a magnetic field removes the degeneracy of the nuclei spin states. Thus, spin "1/2" nuclei will split into two different populations of energy levels, one parallel and the other one anti-parallel to the external magnetic field (Figure 1.8).

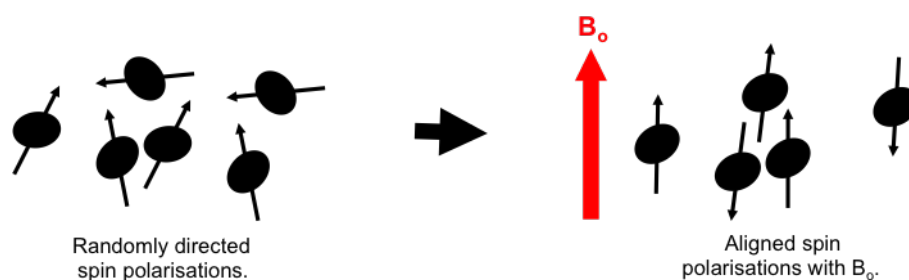


Figure 1.8: Spin polarisation. These schematic shows an isotropic distribution of the spin polarisation in the absence of an external field (B_0). As soon as the B_0 is applied the spin polarisation become aligned along B_0 . The magnetic moment will either be aligned with B_0 (for nuclei with $\gamma \geq 0$), or anti-parallel to B_0 (for nuclei with $\gamma \leq 0$).

This splitting is specific to each nucleus, it depends on the chemical environment and is proportional to the external magnetic field direction. The distribution among the nuclear spin energy levels is determined (as always) by the Boltzmann distribution (5.1), where N_{upper} and N_{lower} represent the population number of nuclei in upper and lower energy states, respectively, k is the Boltzmann constant, and T is the absolute temperature in kelvin (Figure 1.9). This equation reveals that both energy levels are almost equally populated. As a result, NMR is an insensitive form of spectroscopy. However, the small excess number of nuclei in the lower energy state is sufficient to

measure a net macroscopic magnetisation.

$$\frac{N_{\text{upper}}}{N_{\text{lower}}} = e^{-\frac{h\gamma B}{2kT}} \quad (1.3)$$

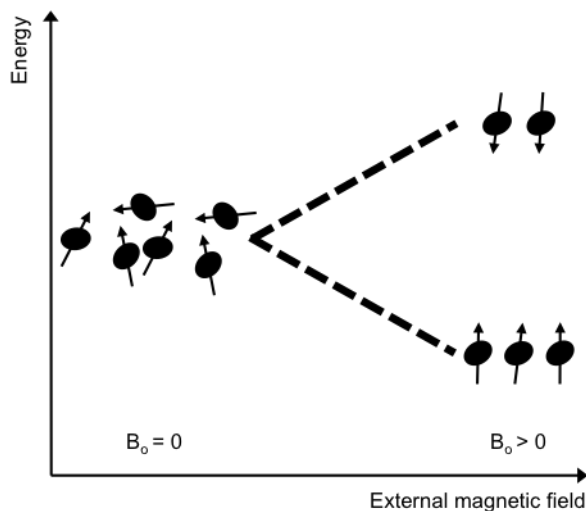


Figure 1.9: The Zeeman effect. The Zeeman Effect is the name given to the splitting of the energy levels by B_0 for a nuclei with $I=1/2$.

1.5.2 Free Precession.

When a particle with a magnetic moment is placed in an external magnetic field (B_0), it experiences a torque and precess around the field at a frequency known as the *Larmour Frequency* ω_0 . The *Larmour frequency* is identical to the resonance frequency of the nucleus as shown in Equation (1.4), and depends on the constant known as the *gyromagnetic ratio*, γ and the magnetic field applied B_0 . Each nuclei possesses a unique γ value.

$$\omega_0 = -\gamma B_0 \quad (1.4)$$

The behaviour is comparable to the action of a gyroscope in a gravitational field. Unlike the gyroscope, the spin energy is quantised and thus the angle and

frequency of precession are fixed. We previously said that in the magnetic field B_0 , the $+z$ component of the magnetic moment of a spin $-1/2$ particle aligns parallel to or anti-parallel to the magnetic field direction. In Figure 1.10 the vector model used to represent electron spins is utilised to represent the x and y component of nuclear magnetic moment.

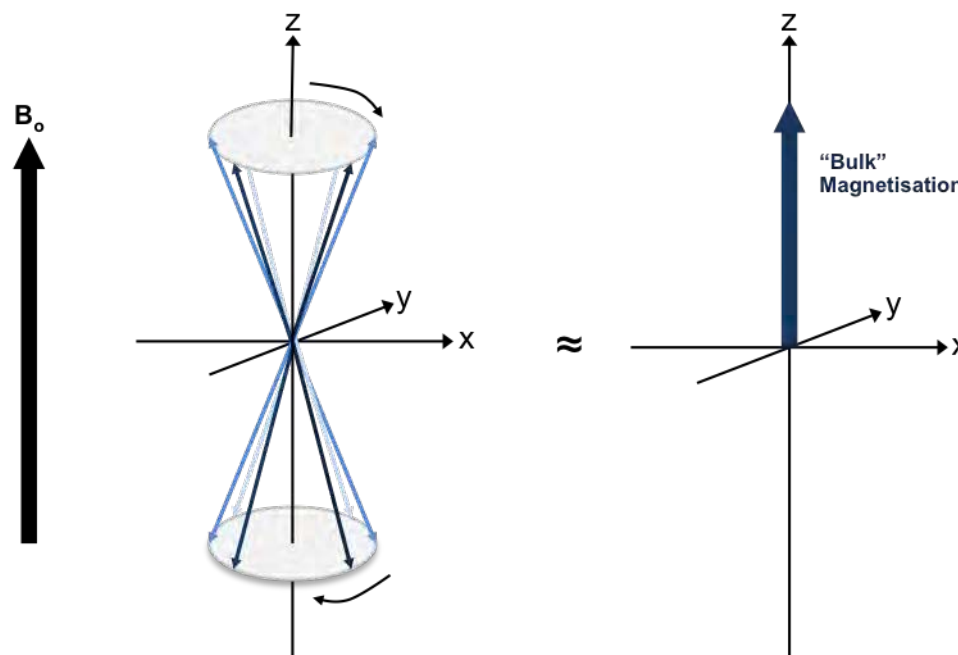


Figure 1.10: Free precession. The two spin states of the nucleus are represented as before (parallel and anti-parallel to B_0), but now the Larmour precession is included. Since there is a slight excess of spin in the ground state (parallel spins to B_0), there is a bulk magnetisation along the z -axis.

In a real sample (imagining all spins to be identical for now), there are slightly more spins in the "spin up" configuration (lower energy state according to Boltzmann's Equation (5.1)). The phases of the individual spins are random. Thus there is an overall magnetisation along the z direction (i.e. along the static field B_0 direction), but no net magnetisation perpendicular to the field because on average the magnetisation distribution at thermal equilibrium is cylindrically symmetrical around the z -axis (Levitt 2001). Instead of measuring the nuclear spin magnetisation along the field (i.e. z -axis), NMR spectroscopy measures the

magnetisation perpendicular to the field. Application of a radio frequency pulse (RF pulse) along the xy -plane causes the bulk magnetisation to precess about it at a frequency related to the RF field (Figure 1.11).

The magnetisation precesses in the xy -plane. By varying the length of time for which the RF field is applied, we can alter the angle through which the magnetisation is rotated. The most common pulse is the 90° excitation pulse. The effect of this pulse is to cause the spins to cluster, producing a net magnetisation along the x -axis in the sample, generating a transverse magnetisation.

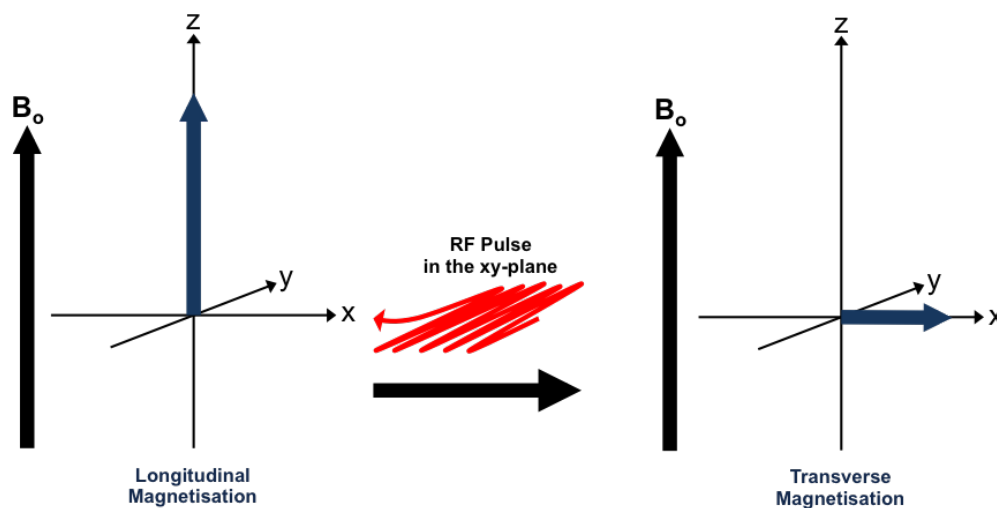


Figure 1.11: Radio frequency pulse. The RF pulse is applied perpendicular to the external field (B_0) along the xy -plane.

1.5.3 NMR Signal.

After a 90° pulse, the transverse magnetisation will return back to its thermal equilibrium position along the z -axis. This is due to T_1 and T_2 , known as the longitudinal and transverse relaxation times respectively. T_1 follows the return to the Boltzmann equilibrium population (i.e. to the initial z direction) of spin-states through spin-lattice relaxation. Additionally, in an actual sample the

precise Larmour frequency of spins will vary due to changing local environments (i.e. neighbouring groups, orientation of the molecule with respect to the axis of the applied field B_0). This means that in the xy-plane some spins precess about z-axis faster (or slower) than the central radio frequency, as a result the spins will gradually fan out in the xy-plane and the coherence is eventually lost (Figure 1.12). This process is called the spin-spin relaxation process, which removes coherence with a time constant T_2 . Individual spins with different resonance frequencies due to chemical shift or coupling precess about the xy-plane at different frequencies relative to the RF pulse (i.e. B_1). This precession produces a radio frequency field in the detector coils. Since the NMR spectrometer detects only the transverse magnetisation, the signal at the coils is a decay curve with frequency components for each different resonance frequency (Figure 1.12). The decay and oscillation of the transverse magnetisation over time can be measured and is called the NMR signal or free induction decay (*FID*). Using the mathematical technique of Fourier Transformation, this FID spectrum can be transformed into the frequency spectrum — the usual NMR spectrum (Figure 1.12).

Figure 1.12: NMR signal. The precession and relaxation of the transverse magnetisation back to longitudinal magnetisation is represented. The signal is detected as function of time along the xy-plane (*FID*). The signal is then Fourier transformed to obtain the NMR spectrum.

1.5.4 Nuclear Shielding.

As seen above, the resonance NMR frequency of a nucleus is given by Equation (1.4). The local magnetic field experienced by a nucleus in an atom differs from

that of a bare nucleus. In an atom, the external field (B_0) causes circulation of electrons within their atomic orbitals. The circular movement of charged particles such as electrons induces a current and as a result a small magnetic field B' is established in the opposite direction to the applied field. Subsequently the nucleus is shielded from the external field applied by the surrounding electrons and experiences a reduced external field (i.e. $B=B_0-B'$). The field at the nucleus may be written as (1.5).

$$B = B_0 - B' = B_0(1 - \sigma) \quad (1.5)$$

where σ is the shielding constant

Electronegative atoms (e.g. oxygen) are electron-withdrawing atoms that will increase the local-field of the neighbouring nucleus. Thus, the nucleus will be deshielded and appear at higher resonance frequency than expected, downfield in the spectrum (Figure 1.13). However, the electron movement is more complex in molecules. Electron currents provided by neighbouring groups (e.g. delocalisation of electrons in an aromatic group produces a *ring-current* effect) can also lead to shielding or deshielding of nuclei. Thus the resonant frequency of a nucleus depends also on its local environment. Accordingly, different nuclei in a molecule will resonate at slightly different frequencies causing each nucleus to precess at a different Larmour frequency creating different resonance frequencies (chemical shift) in the NMR spectrum.

Thus, the chemical shift will be specific to a functional group (i.e. aliphatic protons) and enable the identification of secondary structure in proteins. Indeed,

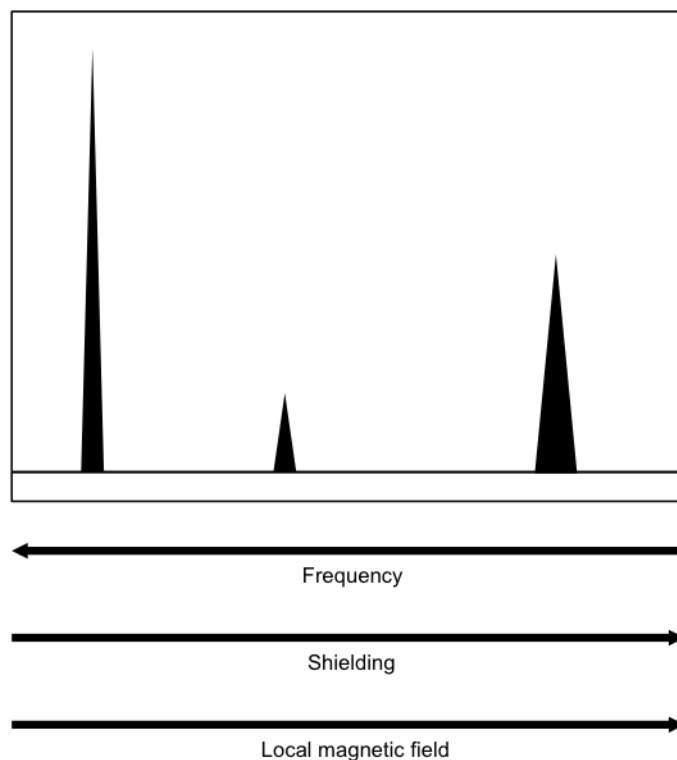


Figure 1.13: Nuclear shielding. 1D ^1H NMR spectrum showing where chemical shifts of more or less shielded nuclei appear on the spectra, along with how it relates to the magnetic field generated around the nuclei.

a rapid 1D proton NMR spectrum can help determine whether the protein is folded and therefore likely to be functional for further analyses. The idea is that if a protein is well folded, the peaks in the 1D NMR spectrum are sharp and narrow and cover a large range of chemical shifts, whereas, if the protein is unfolded or partially folded, the peaks are broader and not as widely dispersed due to loss of different local environments. A spectrum with good dispersion, i.e. peaks spanning a wide ppm range, is a good indication of a folded protein. Unfortunately, 1D proton NMR spectra of proteins can get very complicated. Each peak corresponds to a particular proton in the protein. So that means CPG2, a protein of ≈ 400 amino acids, will have >400 peaks ^1H peaks, many of them overlapping.

To overcome problems caused by overlap in the proton spectrum, protein

NMR is usually multidimensional (2D, 3D, or even 4D). The extremely overlapping 1D NMR spectrum is spread into a carbon or nitrogen dimension. 2D NMR is used to determine which nuclear spins are coupled (through space or bond) to one another. An example is Heteronuclear Single Quantum Coherence (*HSQC*), an heteronuclear experiment which correlates the chemical shifts of protons and ^{15}N -nuclei that are directly bound. As a result, the proton resonances are scattered according to the chemical shifts (discrete spots) of ^{15}N to which they are coupled.

1.5.5 The Notorious Size-Limitation.

The primary limitation in protein solution NMR is the molecular weight restriction. Even though one of the largest structures nearly determined by NMR is an 84 kDa enzyme known as the malate synthase G (Tugarinov et al. 2005), the actual limit for routine studies is in the region of twenty to thirty kilodaltons (Keeler 2011). Larger proteins yield broad NMR signals that are eventually non observable. Two main problems occur as the protein size increases. First, the line width of an NMR signal at half width $\Delta \nu$ is related to the transverse relaxation time (T_2) as described in Equation (1.6) (Bieri et al. 2011).

$$\Delta \nu = \frac{1}{\pi T_2} \tag{1.6}$$

T_2 in most molecules is caused by the external magnetic field variation modulated by the molecular motion of the protein. The molecular motion can be defined by the rotational correlation time for the molecule τ_c , which is the

average time taken by the protein to end up at an orientation about one radian from its initial position (assuming it behaves more or less spherically) (Keeler 2011). The rotational correlation time can be estimated using Equation (1.7).

$$\tau_c = \frac{MV}{\eta RT} \quad (1.7)$$

Where M is the molecular mass, V the molecular volume, η the viscosity and T the temperature (Kay 2005). Thus the tumbling rate is *per se* proportional to the molecular size and shape. As a consequence, slower tumbling rates of huge biomolecules increase transverse relaxation rates, and broaden NMR signals. This gives rise to the second problem, which is the superposition of many NMR signals as peak width increases causing ambiguous assignments. Because of considerable relaxations, peaks can also disappear. Nevertheless, this limitation is constantly driven back by progress made in experimental NMR techniques and development of novel strategies for isotopic labelling of proteins, such as TROSY and perdeuteration respectively.

1.5.6 Transverse Relaxation-Optimised Spectroscopy (TROSY).

Multidimensional NMR spectroscopy provides much better resolution with fewer signals cluttering the spectrum. Unfortunately, isotopic labelling and standard triple-resonance approaches cannot extend application of NMR techniques above the size limit of thirty kilodaltons (Matthews 2004). The transverse relaxation-

optimised spectroscopy (TROSY) methods can extend this range to higher mass (Pervushin et al. 1997).

As the name suggests, it exploits attenuation of the transverse signal relaxation. Transverse signals are the observable signals in NMR spectrum, but their relaxation leads to fast exponential signal decay. As mentioned earlier, transverse relaxation rates are proportional to the molecular weight of a protein. Relaxation for larger proteins at a high magnetic field strength is dominated by dipole-dipole relaxation (DD) and chemical shift anisotropy (CSA) mechanisms. DD relaxation are due to the motion of the nearby magnetic dipole moment (μ) especially with protons, while CSA arises due to various orientation of the molecule with respect to the magnetic field B_0 (Levitt 2001). In solution NMR spectroscopy, CSA is averaged out by molecular tumbling, but the latter is significantly reduced in large proteins. The interference between the two DD and CSA relaxation mechanisms can be subtractive, resulting in slower relaxation of the observable transverse signal, which can be carefully retained and observed in TROSY type experiments. Pervushin et al. (1997) demonstrated that TROSY produced identical information to a ^1H - ^{15}N HSQC (the resulting spectrum is two-dimensional with one axis for ^1H and the other for the ^{15}N), but with much higher resolution for deuterated protein samples because protons are the highest source of DD relaxation (Pervushin et al. 2000).

1.6 NMR Spectroscopy in the Determination of the Binding Interface.

There are times where high-resolution structure determination is not achievable (i.e. study of large proteins). What follows is a description of the NMR techniques that can be used to study the molecular changes associated with ligand-binding interactions at an atomistic level. The effects of ligand binding can be studied via two approaches: ligand-based and protein-based methods.

1.6.1 Saturation Transfer: a Ligand-Based Approach.

In ligand-based experiments, isotopic labelling of the target protein is unnecessary, since only ligand signals are monitored. To study molecular interactions in solution, the Nuclear Overhauser Effect (*NOE*) can be exploited. *NOE* may be described as the change in intensity of the NMR signal when the spin transitions of another nucleus are perturbed from their equilibrium population by radio-frequency saturation (Figure 1.14) (Bieri et al. 2011). This dipole-dipole relaxation has a $1/r^6$ distance dependence, hence only nuclei that are close in space can experience the *NOE* effect.

In the present study, when a ligand binds to a protein, the *NOEs* undergo extreme changes leading to the observation of transferred *NOE*. When a ligand is bound to the enzyme, it acts as if it is part of a large molecule and adopts the corresponding *NOE* behaviour, an intense negative *NOE* called transferred *NOE*. Saturation transfer difference (*STD*) NMR is based on the *NOE* principle

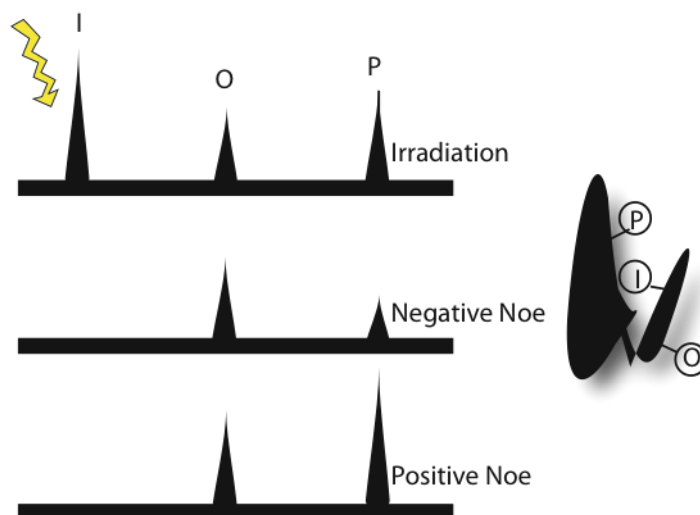


Figure 1.14: NOE effect. When an electromagnetic radiation is applied on one resonance of spin I , a rise (positive NOE) or a drop (negative NOE) in intensity of other NMR signals such as spin P is observed when spin P is close in space to spin I . Small molecules have positive NOEs, whereas bulky molecules have negative NOEs due to tumbling times (Williamson 2009).

(Angulo and Nieto 2011). It is used to identify the elements of the ligand that are in close contact with the protein upon binding. STD-NMR can be described as a three step approach (Figure 1.15). First, an NMR spectrum without protein irradiation is recorded, resulting in the normal spectrum of the free ligand being observed (off-resonance spectrum). Then, a resonance of the protein, where no ligand signals are present, is irradiated leading to a selective and efficient intramolecular saturation of the protein by spin diffusion. This saturation transfers the irradiation to the areas of the ligand bound to the protein. By exchange (free / bound state) that saturation is carried by the ligand into solution where it is detected (on-resonance spectrum). The delivered saturation to the individual proton signals of a ligand reflects the proximity of these to the protein surface. Finally the subtraction of the two spectra (on-resonance minus off-resonance) results in a difference spectrum.

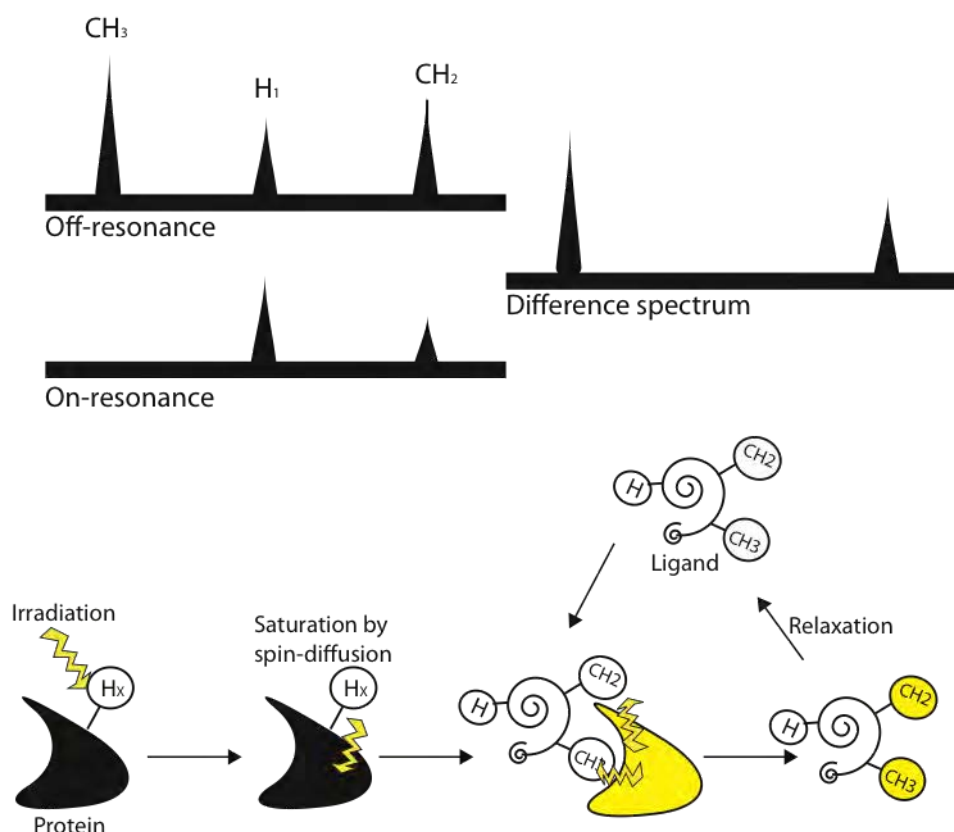


Figure 1.15: Saturation transfer difference NMR spectroscopy.. First a non-saturated 1D ^1H NMR spectrum is recorded, resulting in the spectrum of the free ligand in solution. In a second step, one resonance of the protein, where no ligand resonance frequencies are present, is saturated. This leads to saturation of the rest of the protein due spin diffusion. Upon binding of the ligand the saturation is transferred to protons of the ligand that are in close contact with the protein. By exchange between free and bound states, the saturated ligand is transferred into solution, where the saturation is detected (on-resonance). The subtraction of the on-resonance spectrum from the off-resonance spectrum results in the saturation transfer difference spectrum, where we can observe directly saturated ligand signals by the protein which would have experienced a decrease in the intensity of their NMR signal. (Heller and Kessler 2001).

1.6.2 Chemical Shift Perturbation: a Protein-based Approach.

The major aim of investigating protein interaction with molecules is the characterisation of the interactions sites on both components. The most commonly used NMR experiment for mapping ligand interactions of the surface of the protein is the chemical shift perturbation resulting from complex formation. Since we will be monitoring the protein NMR resonances, the ligand

can be of any type as long as the NMR resonances of the target protein are easily detectable. Because chemical shifts are affected by the changes in the local electronic environment, binding can be detected by comparing the HSQC spectra of the protein in the absence and presence of the ligand.

HSQC displays all ^1H - ^{15}N correlations present within the target protein. For the most part, these are the backbone residues of all non-proline amino acids present within the primary sequence of the protein. However, the sidechains of tryptophan, arginine, lysine, asparagine and glutamine residues, and also histidine (depending on the protonation state of the imidazole ring), can also be visible. As a result, the HSQC experiment provides a "fingerprint" of a protein structure likely to succeed because each protein has a unique pattern of N-H correlations for each residues. However, it necessitates the production of recombinant protein with ^{15}N labelling as a minimal prerequisite. For large proteins, a particular labelling strategy (i.e. perdeuteration) in combination with the appropriate NMR experiment (i.e TROSY) may be required to increase sensitivity and resolution in the HSQC spectrum. Still this type of study usually requires a crystal structure of the target protein, but also an extensively assigned HSQC, in order to map the chemical shift perturbation onto the structure and locate the ligand-protein binding site onto a three dimensional structure of the protein.

Not only can HSQC experiment be used to pinpoint the interaction site of the ligand in the protein, it can also estimate the kinetic characteristics of the interaction. The easiest approach to determine the binding constant is to titrate increasing concentration of the ligand into the protein solution. The kinetics

of binding can then be classified by the overall chemical shift changes observed within the HSQC spectrum during the titration. If the protein-ligand complex dissociation constant is fast on the chemical shift and relaxation rate timescales (i.e. weak ligand-protein binding), one coalesced peak (two peaks have become one) at a weighted averaged chemical shift between free and bound protein is observed throughout the titration (Figure 1.16-A) (Williamson 2013). In cases of slow exchange (usually associated with "strong" binding) separate signals are observed for the free and bound forms of the protein (Williamson 2013).

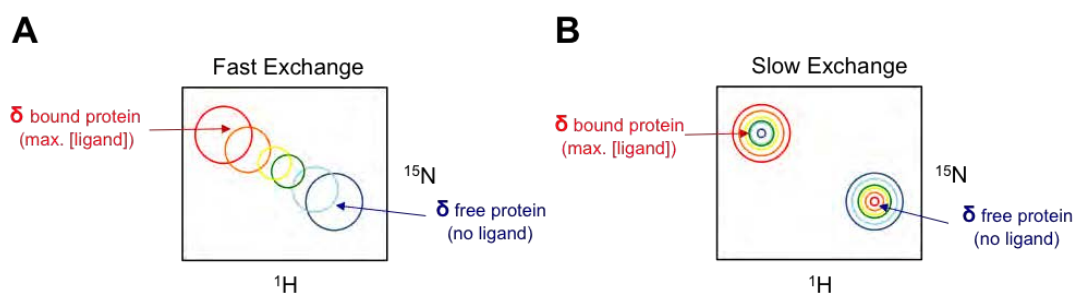


Figure 1.16: Exchange regime. A) In fast exchange, upon titration, the peaks will move smoothly from the chemical shift resonance of the free protein (blue) to the chemical shift resonance of the bound protein; B) In slow exchange, the signal of the free protein (blue) will gradually decrease in intensity as the signal of the bound protein increases close to ligand saturation (red). Adapted from Williamson (2013).

1.7 Research Aims.

The aim of this work was to further our understanding of the mode of binding of peptido-mimetic ligands to CPG2 by acquiring structural information of the free enzyme and inhibitor-enzyme complex. Specifically, we focused on characterising the different molecular interactions, occurring between CPG2 and its different binding inhibitor partners. Using a wide range of biophysical methods, we have

tried to determine the interactions involved in CPG2-inhibitor complex formation. The information obtained could then be exploited towards the development of third-generation inhibitors.

Chapter 2

Materials and Methods

2.1 Materials.

2.1.1 Chemicals and reagents.

All chemicals and reagents were purchased from Sigma-Aldrich Company Ltd. (Dorset, UK), Fisher Scientific UK Ltd. (Loughborough, UK), Melford Laboratories Ltd. (Ipswich, UK), New England Biolabs Ltd. (Hitchin, UK), Promega Corporation (Madison, USA), VWR International Ltd. (Lutterworth, UK), Life Technologies Ltd. (Paisley, UK), GE Healthcare (Little Chalfont, UK) unless otherwise stated. Media used for bacterial growth, including distilled water and glass vessels, was sterilised by autoclave or filtered beforehand. All other consumables (i.e. tips) were sterilised prior to use.

2.1.2 Plasmid Constructs and Bacterial Strains.

Plasmids and *Escherichia coli* strains used in this study are illustrated in Tables 2.1 and 2.2, respectively.

Table 2.1: Plasmid constructs used in this study.

Plasmid	Description	Reference	Source
pET28a	Bacterial expression vector with a T7lac promoter, full length CPG2 inserts sequence - P06621 (CBPG_PSES6), C-terminal Histidine tag, <i>Kan^r</i>	Novagen	Mologic Ltd.
pET151-D-TOPO	Bacterial expression vector with a T7lac promoter, mature CPG2 insert sequence - P06621 (CBPG_PSES6), N-terminal Histidine tag with a TEV cleavage site, <i>Amp^r</i>	Invitrogen	Invitrogen
pRK793	Bacterial expression vector with a T7lac promoter, Tobacco Etch Virus (<i>TEV</i>) insert sequence with a N-terminal Histidine tag, <i>Amp^r</i>	Kapust et al. (2001)	Challis Group

Table 2.2: *E. coli* strains used in this study.

Strains	Genotype	Resistance to phage	Reference
DH5 α	supE44, Δ lacU169(ϕ 80lacZ Δ M15), hsdR17, redA1, endA1, gyrA96, this-1, relA1. DH5 α is frequently used for routine cloning applications.	✗	Hanahan (1983)
DH5 α	fhuA2 Δ (argF-lacZ)U169 phoA glnV44 Φ 80 Δ (lacZ)M15 gyrA96 recA1 relA1 thi-1 hsdR17. DH5 α derivative resistant to phage T1 (<i>fhuA2</i>).	✓	Studier and Moffatt (1986)
C41(DE3)	F- ompT hsdSB (rB- mB-) gal dcm (DE3). BL21(DE3) derivative that has at least one uncharacterised mutation preventing cell death associated with expression of many toxic recombinant proteins such as membrane proteins and some cytoplasmic proteins.	✗	Miroux and Walker (1996)
C43(DE3)	F- ompT hsdSB (rB- mB-) gal dcm (DE3). C41(DE3) derivative selecting resistance for a different toxic protein, and transformed with a plasmid verification vector pAVD10. pAVD10 contains the uncF gene encoding the β -subunit of <i>E. coli</i> ATPase tolerated by C43(DE3).	✗	Miroux and Walker (1996)
BL21 (DE3)	F- ompT gal dcm lon hsdSB(rB- mB-) λ (DE3) [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]. BL21 (DE3) contains the λ DE3 lysogen expressing the T7 RNA polymerase from the lacUV5 promoter by IPTG induction.	✗	Miroux and Walker (1996)
BL21(DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] hsdS, DE3 = sBamHI ϕ Δ EcoRI-B int:::(lacI::PlacUV5::T7 gene1) i21 nin5. BL21(DE3) derivative resistant to phage T1 (<i>fhuA2</i>).	✓	Miroux and Walker (1996)
BL21-CodonPlus (DE3)-RIL	F- ompT hsdS(rB- mB-) dcm+ Tetr <i>E.coli</i> (DE3)-RIL gal λ (DE3) endA Hte [argU ileY leuW <i>Camr</i>]. These cells contain extra copies of the argU, ileY, and leuW tRNA genes to overcome difference in codon usage between different organisms.	✗	Miroux and Walker (1996)

2.1.3 Primers.

Table 2.3 lists all the primers used in this study, which were supplied by Sigma (Dorset, UK), unless stated otherwise. A codon-optimised version of the full-length CPG2 gene from *Pseudomonas sp* strain RS16 in the pET28a plasmid (Novagen) was obtained from Mologic Ltd. and used as a template to amplify the mature wild-type CPG2 gene in the absence of the signal peptide (CPG₂₂₃₋₄₁₅, residues 23-415), as well as the isolated non-contiguous catalytic domain (CPG2_{CAT}, composed of residues 23-214 and residues 323-415 linked via a non-native alanine) and dimerisation domain (CPG2_{DIM}, residues 213-322), using the polymerase chain reaction (*PCR*). The 5'-sense primers encoded a CACC site, and the 3'-antisense primers were designed to add a stop codon for translation termination. The amplified gene segments were ligated into the pET-151-D-Topo vector (Invitrogen, UK). The rest of the primers were used in site-directed mutagenesis to change particular base pairs in CPG2₂₃₋₄₁₅ gene sequence to study the effect of a mutation on protein activity and stability. The approach used was adapted from the QuikChange® Site-Directed Mutagenesis Kit purchased from Invitrogen (Agilent Technologies, UK).

Table 2.3: Primers used in this study.

Construct Name	Protein	Forward Primer (5' → 3')	Reverse Primer (5' → 3')
CPG2 ₂₂₃₋₄₁₅	protein without leader sequence	CACCGCACTGGCACAGAA	TTATTTACCTGCACCCAG
CPG2 _{DIM}	dimerisation domain	GTTATTGTTACCCGTTAA	TTAACGGGTAAACAATAAC
CPG2 _{CAT}	catalytic domain	CACCGCACTGGCACAGAAA	CGCGCTGGTGCCCGAGGCT
		GCGGGTCGTCTCTGCATTT	TTATTTACCTGCACCCAG
CPG2 _{CAT}	short catalytic domain	CACCGCACTGGCACAGAAA	TTAGCTGGTGCCCGAGGCT
H112A	CPG2 with alanine mutation at position 112	CTGCTGCTGATGAGCGCGATGGATACCGTTTATC	GATAACGGTATCCATCGCGCTCATCAGCAGCAG
D141A	CPG2 with alanine mutation at position 141	CGGTCCGGGTATTGCAGCGGATAAAGTGGTAATG	CATTACCACTTTATCCGCTGCAATACCCGGACCG
E175A	CPG2 with alanine mutation at position 175	CGTTCTGTTTAAATACCGATGCGGAAAAAGGTAGCTTTGGTAG	CTACCAAAGCTACCTTTTCCGCGCATCGGTATTAAACAGAACG
E176A	CPG2 with alanine mutation at position 176	CTGTTTAAATACCGATGAAGCGAAAGGTAGCTTTGGTAGCC	GGCTACCAAAGCTACCTTTTCGGCTTCATCGGTATTAAACAG
E200A	CPG2 with alanine mutation at position 200	GATTATGTTTCTGAGCTTTGCGCGACCAAGTGCCGGTGATG	CATCACCGGCACTGGTCGGCGCAAAAGCTCAGAACATAATC
K208A	CPG2 with alanine mutation at position 408	GTGCCGGTGATGAAGCGCTGAGCCTGGGCACC	GGTGCCCGAGGCTCAGCGCTTCATCACCGGCAC
S210A	CPG2 with alanine mutation at position 210	GCCGGTGATGAAAAAAGTGGCGCTGGGCACCAAGCGGTATTG	GCCGGTGATGAAAAAAGTGGCGCTGGGCACCAAGCGGTATTG
H385A	CPG2 with alanine mutation at position 385	GTCTGCCCTGGTTTTTGGTTATGCCAGCGATAAAGCCGAATATG	CATATTCCGGCTTTTATCGCTGGCATAACCAAAACCAAGGCAGAC
T357A	CPG2 with alanine mutation at position 357	GGTGTTGAAGAAACGTGCGGGTGGTGGCACCGATG	CATCGTGGCCACCACCGCGACGTTCTTCAACACC
G360A1	CPG2 with alanine mutation at position 360	CGTACCGGTGGTGCACCGCATGCAGCA	TGCTGCATCGGTGCGACCAACCGGTACG
G360A2	CPG2 with alanine mutation at position 360	CGTACCGGTGGTGGCACCGCATGCAGCATA	TATGCTGCATCGGTGCGCAACCAACCGGTACG

2.1.4 Antibodies.

The monoclonal mouse anti-N-terminal-polyhistidine (dilution 1 in 3000) and anti-Mouse IgG (Fab specific)-Peroxidase antibody (dilution 1 in 5000) were used in this study and purchased from Sigma (Dorset, UK).

2.2 Methods.

2.2.1 Cloning and DNA Conservation.

Polymerase Chain Reaction, PCR.

The amplification of DNA was carried out using the specific enzymes for each PCR according to the manufacturer's guide. Phusion DNA polymerase (NEB, UK) was used to amplify truncated CPG2 genes and clone into pET-151-D-Topo vector (Invitrogen, UK). PCR amplification was usually carried out using a 0.5 μ L DNA template (0.52 ng), 10 μ L of 5 \times Phusion HF Buffer, 1 μ L of 10 mM of dNTPs mix, 2.5 μ L of 10 mM forward and reverse primers from table 2.3, 0.5 μ L of Phusion DNA polymerase (NEB, UK) and distilled H₂O to a final volume of 50 μ L. The reaction mixtures were briefly centrifuged at 1,500 \times g, before heating in a A Biometer T3 thermocycler. The general cycling conditions used for PCR reactions involving amplification and site-directed mutagenesis are described in Table 2.4.

Table 2.4: Cycling conditions used for PCR amplification.

Cycle Step	Temperature	Time	Cycles
Initial Denaturation	98°C	30 s	1
Denaturation	98°C	5 s	
Annealing [†]	56°C	10 s	30
Extension	72°C	45 s	
Final Extension	72°C	10 min	1
	4°C	On hold	

[†] For PCR reactions, for primers greater than 20 nucleotides in length the annealing step was performed at 3°C above the lowest T_m of the primer pair. When primers with $T_m \geq 72^\circ\text{C}$ were used, a two-step thermo-cycling without a separate annealing step was used.

Agarose Gel Electrophoresis.

The results of most PCR experiments were checked by running a portion of the amplified reaction mixture in an agarose gel. The required amount of agarose (1%, w/v) was added to Tris-acetate-EDTA (*TAE*) buffer (Sanbrook et al. 1989) and dissolved by heating to boiling using a microwave. Safeview Nucleic Acid Stain was added (0.1%, w/v) and, once the gel was set, it was submerged in *TAE* buffer. DNA samples from PCR reactions were prepared by mixing with 2 μL loading dye. Electrophoresis was carried out at 50 volts for 1.5 hours or until the dye front was at least half way through the gel. The gel image was recorded on a gel documentation system. The fragment of interest was excised from the gel under a UV trans-illuminator using a sharp blade. A QIAprep gel extraction kit (Qiagen, Germany) was used to remove the agarose from the band of interest. The DNA concentration was determined using the NanoDrop 2000c UV-Vis spectrophotometer.

Ligation of Linear DNA fragments.

PCR product bands 1 and 2 for CPG2_{CAT} construct, composed of two DNA fragments encoding for residues 23-214 and residues 323-415 respectively were excised and purified using a QIAquick Gel Extraction Kit after confirmation by visualisation on an agarose gel (see previous section). These two fragments were ligated at a nanogram ratio concentration of 3:1 PCR product 1: PCR product 2 in the presence of 1 U of T4 DNA ligase polymerase (Invitrogen, UK) according to manufacturer's instruction in a volume of 20 μ L. Reaction mixtures were incubated at 16°C overnight. The next day, the polymerase were heat inactivated at 65°C for 10 minutes before cloning in expression vector pET151-D-TOPO (Invitrogen, UK) (see next Section).

Cloning procedure.

Amplified PCR products of truncated CPG2 gene were inserted in the *E.coli* expression vector pET151-D-TOPO (Invitrogen, UK). The pET Directional TOPO expression kit utilises a 5-minute cloning strategy to directionally clone a blunt-end PCR product into a vector with no ligase, post-PCR procedures or restriction enzymes required. In this system, the PCR product is directionally cloned by adding four bases to the forward primers CACC (see Table 2.3). The PCR product is stabilised in the correct orientation when the overhang in the cloning vector GTGG anneals to the added bases in the 5' end of the PCR product. All amplified gene segments were ligated into the pET-151-D-TOPO vector with an N-terminal tag containing the V5 epitope and a hexahistidine

affinity tag, which could be removed by TEV cleavage. The cloning reaction mixture was prepared using the reagents described in Table 2.5. The mixture was mixed gently with the end of a pipette tip and incubated at room temperature (22-23°C) for 5 minutes. For optimal results, a 2:1 molar ratio of PCR product to vector was used. Once the cloning reaction was performed, chemically competent *E.coli* strains were transformed with the new pET TOPO constructs (see Section 2.2.2).

Table 2.5: Cloning reaction mixture.

PCR product	1 μ L
Salt solution	1 μ L
Sterile water	3 μ L
Vector	1 μ L
Total volume	6 μ L

Site-Directed Mutagenesis.

Site-specific mutation protocol to substitute a specific amino acid in CPG2₂₃₋₄₁₅ was performed by PCR with use of the QuikChange® Site-Directed Mutagenesis Kit (Agilent Technologies, UK) and the oligonucleotides listed in Table 2.3, with the pET151-D-TOPO CPG2₂₃₋₄₁₅ plasmid as a template. This method allows the production of mutants in a rapid three-step procedure involving 1) mutant strand synthesis with two synthetic oligonucleotide primers (see Table 2.3), both containing the desired mutation and each complementary to opposite strands of the vector using temperature cycling where the PCR reaction goes around the entire vector (see Section 2.2.1), 2) digestion of methylated parent plasmid

strand by using 1 μL of DPN1 restriction enzyme at 37°C for an hour in the PCR mixture to select for mutation-containing synthesised DNA only, and 3) transformation of 15 μL of the final reaction into 100 μL of DH5 α competent cells. The PCR reaction was performed using the high fidelity Phusion DNA polymerase to minimise the chances of introducing unwanted mutations in both the gene and the plasmid DNA backbone. Finally, a colony was picked, the plasmid purified and sequenced to check for correct mutation and any PCR introduced errors.

PCR Screening of Bacterial Colonies.

After ligation and plating, PCR screening was performed to test for positive insertion of the gene of interest within the vector. A well isolated colony was picked and transferred to 50 L of sterile water. Part of the colony was transferred to 10 mL LB media containing the appropriate antibiotic for overnight culture and miniprep. 1 μL of the inoculated solution was added to 9 μL of PCR pre-mix. An appropriate volume of PCR pre-mix was set up as described in Table 2.6.

PCR was performed according to Section 2.2.1-PCR, and the PCR products were analysed using a 1% agarose gel (see Section 2.2.1-Agarose Gel Electrophoresis).

Table 2.6: PCR screening of bacterial colonies.

	1×Pre-Mix (μL)	20×Pre-Mix (μL)
10x Taq buffer (+KCl, -MgCl ₂)	1	20
MgCl ₂ (50 mM)	0.3	6
dNTPs 10 mM	0.1	2
Forward primer	0.1	2
Reverse primer	0.1	2
W1 detergent	0.5	10
Taq polymerase	0.1	2
distilled H ₂ O	6.8	136

Glycerol Stocks.

The plasmid constructs produced were stored after transformation in *E. coli* strains in glycerol stocks for long-term stability. 800 μL of the overnights prepared in Section 2.2.1-PCR Screening of Bacterial Colonies were transferred to 2 mL cryogenic vials with 400 μL 50% (v/v) glycerol solution. The vials were flash frozen in liquid nitrogen and stored at -80°C for future use.

Purification of Plasmid DNA.

Plasmid DNA of positive clones were purified from a 5 mL cell culture in LB (see Section 2.2.1-PCR) using the QIAGEN Mini-Prep Kit (Qiagen, Germany). This kit employs the alkaline lysis method where the cells are broken open using an alkaline solution consisting of a detergent sodium dodecyl sulfate (*SDS*) and a strong base sodium hydroxide. Through a series of steps involving agitation, precipitation, centrifugation, and the removal of supernatant, cellular debris is removed and the plasmid is isolated and purified by its adsorption onto a silica

matrix, which is driven by the presence of high concentrations of chaotropic salts. Manufacturers instructions were used and plasmid DNA was recovered in 30 μ L of distilled H₂O and stored at -20°C.

Sequence Analysis of Plasmid DNA.

Purified plasmid DNA was dispatched to GATC (Germany) and sequence analysis was performed using vector-specific forward and reverse T7 primers synthesised by GATC to assess the correct orientation of the insert into the vector, presence of start and stop codons and no mutations were incorporated by PCR errors.

2.2.2 Bacterial Growth and Conservation.

Preparation of Competent Cells.

3 mL of an overnight culture of *E.coli* in LB was inoculated into 300 mL of fresh LB broth. The culture was incubated at 37°C and shaken (180 rpm). When turbidity was observed, an aliquot was removed and the optical density (OD) checked at 600 nm, using LB (no cells, no antibiotic) as a blank. The OD₆₀₀ was examined every half hour to make sure it was not beyond 0.35. The culture was cooled by placing the conical flask on ice for twenty minutes, followed by aliquoting into 50 mL ice-cold falcon tubes. Cells were pelleted by centrifugation 1500×g at 4°C for 10 min. Supernatant was poured off and pellets were resuspended in 30 mL of ice-cold 100 mM CaCl₂. Pellets were retrieved by centrifugation at 1500×g at 4°C for 10 min. Pellets were

resuspended in 2 mL of ice-cold 10 mM CaCl_2 . At this point, the cells were mixed with glycerol as follows: 300 μL of cells with 50 μL of 50% glycerol solution into 1.5 mL tubes. These eppendorf tubes were stored at -80°C until further use.

Cell Transformation.

Competent cells (50 μL) were transferred with 1 μL of plasmid DNA and incubated on ice for 30 min. The cells were then heat-shocked for two minutes at 42°C , followed by incubation on ice for a further two minutes. LB broth (1 mL) was added to the cell culture and then subjected to shaking (180 rpm) incubation for 1 hour at 37°C . 200 μL of the transformation mix was spread onto LB agar plates containing the appropriate antibiotic (Table 2.7) and incubated inverted overnight at 37°C .

Table 2.7: Recommended antibiotic concentrations for rich and minimal media.

Antibiotic	Abbreviation	Final Concentration ($\mu\text{g}/\text{mL}$)
Kanamycin	Kan	50
Ampicillin	Amp	100
Chloramphenicol	Chl	34

2.2.3 Preparation of CPG2.

2×YT medium.

2×yeast extract and tryptone medium (2×YT) was used to produce protein used for ligand based NMR spectroscopy studies. 2×YT was prepared by adding 16g of tryptone, 10 g of yeast extract and 5 g of NaCl to 800 mL of distilled H₂O. The pH of the media was adjusted to 7.3 using the pH meter. The volume was topped up to 1 L using distilled H₂O.

Preparation of CPG2.

1 L of 2×YT with ampicillin was inoculated with 10 mL of the overnight culture and shaken at 180 rpm and 37°C until the OD₆₀₀ of 2 was reached during the exponential growth phase. Crude pre-induction samples (1 mL) were removed upon an OD₆₀₀ of 2 being attained and kept for SDS-PAGE analysis. Induction of CPG2 expression was achieved with the addition of Isopropyl β -D-1-thiogalactopyranoside (*IPTG*) to a final concentration of 0.5 mM, and the culture was transferred to the 15°C incubator, which is the most favourable temperature for expression (see Chapter 3). Crude post-induction samples (1 mL) were removed at 4, 24 and 48 hours post-induction and OD₆₀₀ readings were taken. Later they were submitted to sonication. The remaining culture was transferred evenly to 500 ml centrifuge tubes and centrifuged at 4,300×g for 20 min at 4°C. The recovered pellets were then resuspended in 10 ml of Tris buffer (20 mM Tris, 137 mM NaCl, 1 mg/mL lysosome, 1 mM 4-(2-Aminoethyl)

benzenesulfonyl fluoride hydrochloride, pH 7.3). If time was an issue, the pelleted cells by centrifugation at $8,300\times g$ were stored at -80°C overnight; otherwise the cell disruptor was used to lyse the cells.

2.2.4 Preparation of Isotopically Labelled CPG2.

Preparation of Minimal Media.

Minimal media (M9) was used to produce isotopically either singly (^{15}N), doubly ($^{15}\text{N}/^{13}\text{C}$) or triply ($^2\text{H}/^{15}\text{N}/^{13}\text{C}$) labelled protein for enzyme-based NMR spectroscopy studies. M9 was prepared by adding 200 mL of $5\times\text{M9}$ salts (50 mM Na_2HPO_4 , 50 mM K_2HPO_4 , 5 mM Na_2SO_4), 1 g of ^{15}N -ammonium chloride, 4 g of ^{13}C -glucose. While stirring to circumvent precipitation, $100\text{ }\mu\text{M}$ and 2 mM of iron chloride and magnesium sulphate, respectively, were added drop wise, followed by 10 mL of $100\times\text{BME}$ vitamin solution. The pH of the solution was adjusted to pH 7.3 and sterile distilled H_2O was added to a final volume of 1 L. Finally, the solution was supplemented with the appropriate antibiotic for plasmid selection. For perdeuterated (100% deuteration) proteins, deuterated water was used instead of water.

Traditional IPTG-Induction Expression in M9.

Labelled M9 medium (1 L) inoculated with cells from the glycerol stock (volume of starter culture corresponds to 10% of the final M9 volume) and shaken at 180 rpm at 37°C for 24 hours. When the OD_{600} of the starter culture reached 1, 1L

of fresh labelled M9 culture was inoculated with the starter culture to obtain an initial OD_{600} of 0.1 and was shaken at 180 rpm at 37°C for growth. As soon as $OD_{600} \approx 1$, cells were induced with 0.5 mM IPTG and incubated at 15°C for 20 hours at 180 rpm (optimal conditions).

2.2.5 Protein Expression.

High-cell Density Method and IPTG-Induced Expression.

This expression method, instead of the traditional IPTG-induction expression in M9, was used to reduce the cost of labelled media. *E. coli* cells were grown in rich medium at 37°C. Once the cell density reached $OD_{600} \approx 2$, the cell culture was switched to labelled M9 media by gently spinning cells down at 1,500×g for 5 min. The pellet was repeatedly washed with M9 to remove rich medium. Cells were re-suspended into the concentrated volume of minimal medium (2×). The cells were cultured for another 1-1.5 h without adding IPTG at 37°C (OD_{600} should increase by 1-2 units). IPTG (0.5 mM) was then added to induce protein expression. Cells were then incubated at 15°C for 20 hours, which was the determined optimal expression time.

Double Colony Selection.

A double colony selection was performed for bacterial expression of triple-labelled protein in D₂O (Murray et al. 2012). For the first colony selection, LB- agar plates were prepared and the glycerol stock was streaked on to it and left overnight in the

incubator at 30°C. The next day, 9 colonies (with the first colony serving as the negative control) were chosen for expression and a master plate was prepared by streaking the 9 colonies onto it. 8 colonies were induced with 0.5 mM IPTG and expressed at 15°C in LB with 50% D₂O and checked for CPG2 level expression after 24 hours via SDS-PAGE and western blot analysis. The colony with the thickest expression band for CPG2 was selected from the master plate and used to prepare 5 mL overnight culture in LB with 100% D₂O at 30°C. This culture was spread onto an LB-agar plate with 50% D₂O and incubated at 37°C overnight. A second round of colony selection was repeated, but this time in LB with 100% D₂O. At the end of the procedure, the colony expressing the thickest protein-band in LB with 100% D₂O was selected and used to prepare the final glycerol stock (800 μ L of overnight cell culture of the selected colony from the second master plate with 400 μ L of 50% glycerol), which was stored at -80°C.

Cell Lysis by Ultrasound Sonication.

Ultrasound sonication was used to disrupt the bacterial cell extracts and lyse cells (5 mL and above) by high frequency sound. Sonicator probes were pre-chilled on ice before use. All samples remained on ice throughout the sonication process. All crude samples (1 mL) and lysed bacterial pellets (10 mL) were subjected to sonication using 6 cycles of 30 seconds on ice followed by 10 seconds of sonication using a 5 mm probe, and 4 cycles of 1 minute on ice followed by 1 minute of sonication using the a 7 mm probe, respectively.

Cell Lysis by Cell disruption.

A cell disruptor was later used to disrupt cell extracts (10 mL and above) at 30 kPSI. In this method, the pressure drops by transferring the cell extract from a chamber at high pressure through an orifice into a chamber at low pressure, ensuring cell lysis. It is fast and efficient but causes significant heating of the sample. For this reason, the collector was always kept at 4°C before usage.

2.2.6 Protein Purification.**CPG2 Purification.**

Recombinant CPG2 was isolated from cell debris by ultrasound sonication or cell disruption. Cell pellets from 1 L culture were thawed on ice and resuspended in 10 mL lysis buffer (20 mM Tris, 137 mM NaCl, no EDTA chelating agent, lysozyme 1 ug/mL, 1 mM AEBSF, pH 8) and incubated at 4°C for 30 min. The cells were lysed and insoluble fraction pelleted by centrifugation (48,000×g, 20 mins, 4°C). The soluble fraction was incubated with 10 mM imidazole and Ni-NTA resin (Novagen, 50% slurry, pre-equilibrated with 10 mM imidazole) for 1 hour. The incubated supernatant-Ni mixture was loaded onto a column and the flow-through collected as the resin set by gravity. The column was washed extensively with 10×bed volumes of Tris buffer (100 mM NaCl, 20 mM Tris, 10 mM imidazole, pH 8). The his-tagged protein was then eluted with 5×bed volumes of elution buffer (100 mM NaCl, 20 mM Tris, 250 mM imidazole, pH 8). The buffer was exchanged against Tris buffer (100 mM NaCl, 20 mM Tris, pH 8)

using PD-10 columns to desalt samples.

His-TEV Protease Expression and Purification.

The 2×YT starter culture (10 mL) was inoculated with cells from glycerol stock and left at 30°C overnight. This outgrowth was then transferred to a final volume of 1 L fresh 2×YT to obtain an initial OD₆₀₀ of 0.1. The cells were then grown at 37°C until the OD₆₀₀ reached ≈ 2 , then it was induced with 0.5 mM IPTG and incubated at 15°C overnight. Recombinant His-TEV protein was then harvested. Cell pellets from 1 L culture were thawed on ice and resuspended in 10 mL Tris lysis buffer (20 mM Tris, 137 mM NaCl, no EDTA — chelating agent, lysozyme 1 μ g/mL, 1 mM AEBSF, pH 8) and incubated at 4°C for 30 min. The cells were lysed and insoluble fraction pelleted by centrifugation (48,000×g, 20 min, 4°C). The soluble fraction was incubated with 10 mM imidazole and Ni-NTA resin (Novagen, 50% slurry, pre-equilibrated with 10 mM imidazole) for 1 hour. The incubated supernatant-Ni mixture was loaded onto a column and the flow-through collected as the resin set by gravity. Column was washed extensively with 10×bed volume of Tris buffer (100 mM NaCl, 20 mM Tris, 10 mM imidazole, pH 8). The his-tagged protein was then eluted with 5×bed volume of elution buffer (Tris buffer, 250 mM imidazole, pH 8) and buffer exchanged against Tris buffer (100 mM NaCl, 20 mM Tris, pH 8) using PD-10 columns to desalt samples.

His-Tag Removal by Enzymatic Cleavage.

Both desalted and purified his-tagged CPG2 and TEV proteins in Tris buffer (100 mM NaCl, 20 mM Tris, pH 8) were incubated together for 48 hours at 4°C for cleavage of the polyhistidine tag.

Purification of the Untagged CPG2.

The overnight CPG2-TEV reaction mixture was incubated with 10 mM imidazole and Ni-NTA resin (Novagen, 50% slurry, pre-equilibrated with 10 mM imidazole) for 1 hour. The incubated supernatant-Ni mixture was loaded onto a column and the flow-through, containing the protein of interest was collected, as the resin set by gravity. The solution volume was slowly reduced using centrifugal units of 20 molecular weight cut-off (Millipore, UK). The unlabelled protein sample was then lyophilised using the freeze drier and stored at -20°C for ligand-based NMR studies, while the labelled protein sample was used straightaway away for protein-based NMR studies.

2.2.7 Quantification of Protein Concentration.

Protein Concentration Determination using UV-Visible Spectrophotometer.

60 μ L of purified CPG2 was transferred to a Thermo Electron Cuvette and the absorbance was read at 280 nm against the blank (100 mM NaCl, 20 mM Tris)

using the UV visible BioMate instrument. The concentration was determined using Beer-Lambert Law Equation (2.1), where A is the absorbance, ϵ the molar absorptivity in $\text{M}^{-1} \text{cm}^{-1}$, l the path length in cm and C the concentration in molar.

$$A = \epsilon l C \quad (2.1)$$

Protein Quantification using BCA Assay.

The bicinchoninic acid (BCA) assay was used for colorimetric detection and quantitation of total protein (Thermo Scientific Pierce, UK). The manufacturer's guidelines were followed to perform the assay. After incubation at room temperature (i.e. 25°C) for 2 hours, the absorbance reading of the sample of interest and the standards at 562 nm were taken using the UV- visible spectrophotometer. Set at 562 nm, the instrument was zeroed using a cuvette filled with water. The average absorbance reading for standard I (i.e. blank) was subtracted from the absorbance reading for each remaining standard and unknown sample. The standard curve was prepared by plotting the average blank-corrected absorbance reading for each standard versus their respective concentrations in $\mu\text{g/mL}$. Using the standard curve, the gradient was be calculated and, using Equation (2.2), the protein concentration was established.

$$\text{Protein Concentration } (\mu\text{g}) = \frac{\text{Absorbance (562 nm)}}{\text{gradient}} \quad (2.2)$$

2.2.8 Protein Detection.

Protein Solubility Assay.

Sonicated crude samples (1 mL) were centrifuged ($5,000\times g$) for 5 min to pellet the insoluble components. The supernatant (soluble fraction) and the pellet (the insoluble fraction) were retained for SDS-PAGE analysis. 10 mL sonicated lysed samples were also centrifuged ($5,000\times g$) for 15 min at 4°C . The supernatant and the pellet were retained for further purification.

Optical Density of Culture Samples.

Sample was diluted to yield an OD_{600} reading between 0.1 and 0.8. Usually a 1:10 dilution of the culture was sufficient. The same media used for growth of the culture was used for zeroing the spectrophotometer and diluting the samples for taking OD_{600} readings. The dilution factor and the OD_{600} reading were recorded to calculate the actual OD_{600} of the culture.

Normalising SDS-PAGE Samples.

To accurately compare the amount of protein present between various culture samples taken at various growth stages, the samples were normalised according to the cell density (OD_{600}) at the time the samples were taken. Maximum volumes that can be loaded into the wells of the SDS-PAGE gel were also taken into account. The volume of samples loaded onto the gels was normalised with respect to the concentration of cells in the pre-induction sample.

SDS-PAGE Gels.

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (*SDS-PAGE*) was used for the separation of proteins based on their molecular weight. All SDS-PAGE gels were prepared using the standard Laemmli Method (Laemmli 1970).

Table 2.8: SDS-PAGE gels sample buffer and running buffer.

(a) Sample Buffer	
SDS	10% w/v
Dithiothreitol	10 mM
Glycerol	20% v/v
Tris-HCl, pH 6.8	0.2 M
Bromophenolblue	0.05% w/v

(b) Running Buffer	
Tris-HCl	25 mM
Glycine	200 mM
SDS	0.1% w/v

The percentage acrylamide was selected based on the molecular weight range of proteins we wished to separate (Table 2.8). A 12% gel was suitable to separate protein in the 10 kDa-200 kDa range (see Table 2.9). Tetramethylethylenediamine (*TEMED*), an essential catalyst for polyacrylamide gel polymerisation, and ammonium persulfate (*APS*), an oxidising agent that is used with *TEMED* to catalyse the polymerisation of acrylamide, were not added until the gel was ready to be poured into the gel casting plates.

Table 2.9: SDS-PAGE gel recipe.**(a)** Resolving Gel Solution (12% acrylamide)

	12%
H ₂ O	10.2 mL
1.5 mM Tris-HCl, pH 8.8	7.5 mL
20% SDS (w/v)	0.15 mL
30% Acrylamide / 0.8% bis-acrylamide (w/v)	12 mL
10% (w/v) ammonium persulfate	0.15 mL
TEMED	0.02 mL

(b) Stacking Gel Solution (4% acrylamide)

	4%
H ₂ O	3.075 mL
0.5 mM Tris-HCl, pH 6.8	1.25 mL
20% SDS (w/v)	0.025 mL
30% Acrylamide / 0.8% bis-acrylamide (w/v)	0.67 mL
10% (w/v) ammonium persulfate	0.025 mL
TEMED	0.005 mL

SDS-PAGE Analysis.

The Mini Protean 3 Electrophoresis System (Bio-Rad) was used to perform all electrophoresis according to the manufacturer's instructions.

1 mL crude samples and insoluble fractions from sonicated samples were centrifuged ($5,000\times g$) for 5 min at 4°C and the supernatant were discarded. Pellets were resuspended in 15 μ L of PBS buffer and 5 μ L Laemmli Buffer. Sonicated soluble samples (1 mL, 10 mL) were prepared for SDS by adding 5 μ L Laemmli Buffer to 15 μ L of sonicated sample. The resulting samples were

heated at 90°C on a heating block for 5 min exactly. The sample was centrifuged (5,000 rpm) and vortexed prior and after heating for 5 min. The volume sample of SDS were normalised with respect to the concentration of cells in the pre-induction sample, and ready for SDS-PAGE analysis. SeeBlue® Plus2 Pre-Stained Standard (Invitrogen, 4-250 kDa) or ColorPlus restained Protein marker Broad range (NEB, 7-175 kDa) were run on gels as protein standards. Gels were run at 125 V for approximately 15 min and further at 200 V for 35 min until the front dye had migrated to the end of the gel. Upon completion, the gel was subjected to Coomassie Blue Staining overnight on a rocker. Gel was then destained with destaining solution (7% acetic acid, 40% ethanol) until the bands were clearly visible and scanned (gels were maintained between 2 plastic sheets) and stored in H₂O.

Coomassie Staining.

Upon completion, the gel was subjected to coomassie blue staining overnight on a rocker. Gels were first immersed in 20 mL of fixing solution [50% (v/v) MeOH, 10% (v/v) acetic acid] for 15 min, then stained in 20 mL coomassie solution [0.025% (w/v) coomassie blue R250, 10% (v/v) MeOH] for 4 to 12 hours. Gels were destained in a 10% (v/v) acetic acid solution until the bands were clearly visible.

Western Blot.

Gels were electro-blotted onto a 0.45 mm pore size nitrocellulose membrane at 100 V, 35 mA for 60 min using the Mini Trans-Blot® Electrophoretic Transfer Cell according to the manufacturer's guide (Bio-Rad).

After the transfer, the membrane was blocked in Tris-Buffered Saline and Tween-20 (*TBST*) (30 mM Tris-HCl, 140 mM NaCl, 3 mM KCl, 0.1 % Tween-20, pH 7.4) and 5% milk solution for maximum 1 hour at room temperature (or overnight at 4°C), then washed with *TBST* buffer 3 times for 5 min. The membrane was inoculated for 1 hour with the primary antibody mouse anti-polyHistidine, which was diluted in *TBST* and 5% milk, and then the washing steps were repeated (3×5 mins in *TBST*). The secondary antibody tablet, anti-mouse anti-polyHistidine-alkaline phosphatase (Sigma), diluted in 10 mL distilled H₂O and used to incubate the blot for 1 hour, followed by 2×30 min washing steps with *TBST*.

Size exclusion size exclusion chromatography analysis.

4 mg/mL of protein solutions (CPG2, CPG2_{CAT}, CPG2_{CAT'} and CPG2_{DIM}) were prepared in buffer A (20 mM Tris, 100 mM NaCl, pH 8). Gel filtration chromatography was performed on a HiLoad 16/600 Superdex 75 g GL column by fast protein liquid chromatography ÄKTA FPLC (GE Healthcare UK Ltd. Buckinghamshire, UK) at 4°C and at a flow rate of 0.5 mL/min. The column was equilibrated with two column volumes of buffer A. The detection was performed by monitoring the column eluate at 280 nm. To determine the

molecular weight, a parallel column was run with the following molecular weight markers at 5 mg/mL concentration each: β -amylase (200 kDa), Albumin (66 kDa) and Cytochrome C (12.4 kDa) (see Figure 2.1).

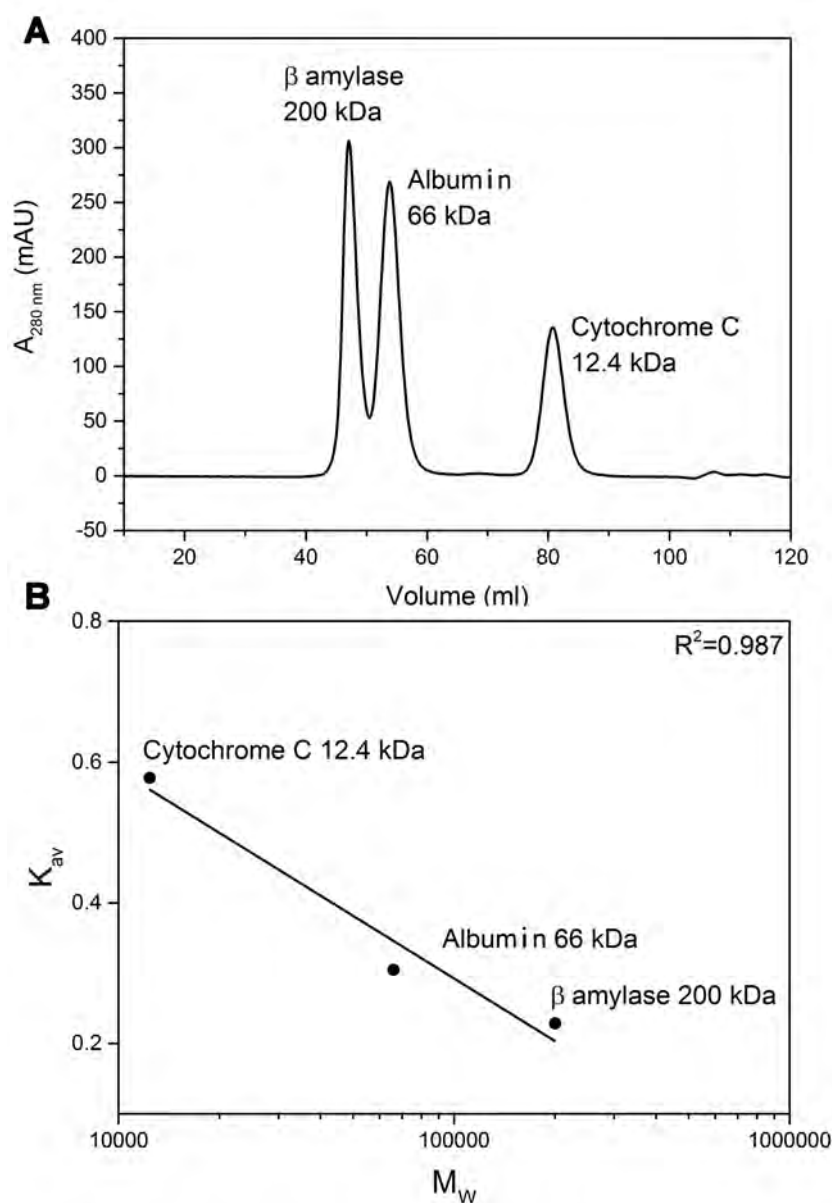


Figure 2.1: Elution and Calibration curve obtained with β -amylase, Albumin and Cytochrome C protein Standards run on HiLoad 16/600 Superdex 75 g GL column.

A) The graph shows elution profile from HiLoad 16/600 Superdex 75 g GL column for calibration standards (β -amylase 200 kDa, Albumen 66 kDa, Cytochrome C 12.4 kDa) as absorption in milli-absorption units (mAU) at 280 nm versus volume in milliliters (mL). An isocratic gradient of 100% of buffer A (20 mM Tris, 100 mM NaCl, pH 8) was used. B) The calibration curve graph represents the gel phase distribution coefficient K_{av} versus the molecular weight (M_w). $K_{av} = (V_e - V_o)/(V_c - V_o)$, where V_e = elution volume, V_o = column void volume, V_c = geometric column volume. Straight line is the calibration curve calculated from the data for molecular weight standards. Using the equation for the line, values for the molecular weights of CPG2, CPG2_{CAT}, CPG2_{CAT}⁺ and CPG2_{DIM} were calculated (see Chapters 2 and 6).

Protein identification by Means of Tryptic Digest and Nano LC-ESI-MS/MS.

Excised bands from coomassie stained gel pieces were processed and tryptically digested using the manufacturer's recommended protocol on the MassPrep robotic protein handling system Facility¹. The extracted peptides from each sample were analysed by means of nano LC-ESI-MS/MS using the NanoAcquity/Ultima Global instrumentation (Waters) using a 30 minutes LC gradient. All MS and MS/MS data were corrected for mass drift using reference data collected from the (Glu1)-Fibrinopeptide B (human-F3261 Sigma) sampled each minute of data collection. A full description of the methodologies employed is available on request. The data were used to interrogate the *E. coli* BL21 database (<http://www.ebi.ac.uk/integr8>) appended with the sequence supplied, using ProteinLynx Global Server v2.5.1.

Identification of proteins via MaXis II

All his-tagged proteins were analysed for their identity under positive-mode using liquid chromatography electrospray ionisation ultra-high resolution quadrupole time-of-flight mass spectrometry (*LC-ESI UHR QTOF*)². Samples at 20 μ M protein in 20 μ M ammonium bicarbonate were diluted by 1 in 5 in 30% acetonitrile (*ACN*) with 0.1% formic acid (*FA*), and 40 μ M aliquots of these were loaded onto a C4 column with 100% water and 0.1% FA as solvent A

¹Dr Suzanne Slade (School of Life Sciences, University of Warwick) performed tryptic digest, protein identification and analysis by Nano LC-ESI-MS/MS.

²Philip Aston (Department of Chemistry, University of Warwick) performed analysis by Bruker MaXis II mass spectrometer.

and 100% ACN with 0.1% FA as solvent B using a 3-step gradient (75% of solvent A for 16 min, 100% of solvent B for 5 min and 75% of solvent A for 11.5 min) and eluted directly into the Bruker MaXis II mass spectrometer. The resulting total ion count (TIC) data (summed intensity across the entire range of masses being detected at every point in the analysis) was extracted and mass averaged in the region containing the proteins of interest, and deconvoluted using the Data Anaysis Software (Bruker Daltonik GmbH; Bremen, Germany).

To confirm the identity of expressed proteins, samples were analysed before TEV digestion (i.e. with the presence of his-tag), by mass spectrometry. Resulting spectra were averaged (see Table 2.10). Full-length his-tagged CPG2₂₃₋₄₁₅, CPG2_{CAT} and CPG2_{DIM} were identified at the exact molecular mass predicted ± 1 Dalton, containing the gene sequence. For CPG2 the $[M+H]^+$ ion was observed at 45477.9 Da (expected molecular ion mass including the his-tag is 45478.6 Da), for CPG2_{CAT} at 33889.7 Da (expected molecular ion mass is 33889.4 Da) and for CPG2_{DIM} at 15649.1 Da (expected molecular ion mass is 15649.7 Da). For his-tagged CPG2_{CAT}, the $[M+H]^+$ ion was observed at 22922.9 Da (expected molecular ion mass is 22580.6 Da) with 342.3 Da of excess mass, which appeared to show post-translational modifications of phosphogluconoylation of His-tag in *E.coli* (258 Da), and a combination of multiple post-translational modifications (i.e. acetylation + 42 Da) at different sites (i.e. different amino acid such as Ser, Thr or Lys) could also explain the remaining 84 Da of excess mass (Cain et al. 2014). The spectra showed some impurities remaining after the first affinity chromatography, but the highest intensity peak was the $[M+H]^+$ peak. All CPG2₂₃₋₄₁₅ mutants were also

identified at the exact molecular mass predicted ± 1 Dalton, except for E175A and E200A, which were at the molecular mass predicted ± 2 Daltons.

Table 2.10: Mass average of all expressed constructs (without TEV protease cleavage) obtained from LC-ESI UHR QTOF mass spectrometry.

Constructs	Calculated Mass (Da)	Observed Mass (Da)	Mass Shift (Da)
CPG2 ₂₃₋₄₁₅	45478.6	45477.9	-0.7
CPG2 _{CAT}	33889.4	33889.7	+0.3
CPG2 _{CAT'}	22580.6	22922.9	+342.3
CPG2 _{DIM}	15649.7	15649.1	-0.6
H112A mutant	45412.6	45411.7	-0.9
H385A mutant	45412.6	45411.8	-0.8
E175A mutant	45420.6	45418.7	-1.9
E200A mutant	45420.6	45418.7	-1.9
D141A mutant	45434.6	45433.7	-0.9
K208A mutant	45421.5	45420.8	-0.7
S210A mutant	45462.6	45461.8	-0.8
T357A mutant	45448.6	45447.7	-0.9

Mass Spectrometry.

A simple protocol for the analysis of intact proteins by matrix-assisted laser-desorption / ionisation time-of-flight mass spectrometry (*MALDI-TOF MS*), utilising a mixture of sinapinic matrix. The degree of isotope incorporation after labelling (i.e. e.g. $^{13}\text{N}/^{13}\text{C}/^2\text{H}$ labelling), his-tag removal by TEV protease and sample homogeneity were also rapidly analysed using MALDI-TOF MS on an Autoflex Speed instrument (Bruker, UK). The protein samples were buffer exchanged into 20 mM ammonium acetate using Amicon

Ultra centrifugal filters with a 20 kDa MWCO (Millipore, UK). One microliter of the protein solution was spotted onto the target MALDI plate and then overlaid with two microliters of the matrix consisting of a saturated solution of sinapinic acid or 2-Cyano-3-(4-hydroxyphenyl) acrylic acid (Sigma, UK) in 50% acetonitrile (Sigma, UK) and 2.5% trifluoroacetic acid (Sigma, UK). The data was acquired with FLeX Control software and processed using FLeX Analysis software by Bruker (Coventry, UK).

2.2.9 Protein Folding Assessment.

Buffer Screen.

Initially, three buffer conditions were tested to extend the stability of CPG₂₂₃₋₄₁₅ and these are summarised in Table 3.3 in Chapter 3. To improve long-term solubility, stability and sample conditions, we applied the hanging drop method used in X-ray crystallography (Ducat et al. 2006). The conditions tested were the pH (5.2, 5.5, 6, 6.5, 7, 7.1, 7.42 7.48, 8.05 and 8.53), the type of salt (NaCl, NaBr, LiCl, KCl, KSCN, Na₂SO₄, and Arg and Glu) and salt concentration (50, 100, 150 and 200 mM). Buffer solutions were prepared at 50 mM with sodium citrate at pH 5, sodium potassium buffer from pH 5 to 7, with Tris buffer from pH 7.5 to 8.5, with HEPES buffer at pH 7, with MOPS buffer pH 7 and with PIPES buffer at pH 7. Each buffer solution was pH adjusted after addition of salt (the final pH of each buffer solution is shown in Table 3.3 in Chapter 3). To determine optimal buffer conditions, the influence of buffer conditions on concentrated protein sample (\approx 23 mg/mL) was determine using the hanging

drop set up with 24 well-tissue culture plates and glass cover slips (Ducruix and Giege 1999). 1 μL aliquots of protein solution and 1 μL of 50 mM reservoir buffer solution were pipette onto the cover slips. The cover slip was inverted and sealed onto the wells using paraffin. The plates were left at room temperature for vapour diffusion to take place. The amount of precipitate were checked after 5 and 11 days by visually examining under the microscope

Spectroscopy Assay for CPG2.

The activity of CPG2 was estimated by measuring its ability to cleave methotrexate (MTX, which absorbs at 320 nm) into two non-absorbing species (Figure 2.2). In the assay, a 50 μL aliquot of CPG2 solution (20 mM Tris-HCl buffer, pH 8) was added to a 1 mL quartz cuvette containing 900 μL of the assay buffer (0.1 M Tris-HCl, 0.2 mM ZnCl_2 , pH 7.5) and 50 μL of 0.6 mM methotrexate solution (0.1 M Tris-HCl, 0.2 mM ZnCl_2 , pH 7.5) pre-equilibrated to 37°C before addition of CPG2. The absorbance at 320 nm was measured using a Perkin Elmer Lambda 35 UV/VIS Spectrometer with Peltier temperature control unit for 1 minute, in timesteps of 2 sec, and absorbance versus time (min) was plotted for each sample. The slope was calculated for the linear portion of the plots. By default, this is defined as occurring between 0 and 1 min. The value of the slope was adjusted to take into account the dilution factor for the enzyme solution and the enzyme activity was obtained in units/mL. One unit is the amount of enzyme required to catalyse the hydrolysis of 1 μmol MTX per minute at 37°C. The absorbance of the assay buffer (blank) was subtracted from all measurements.

To test a potential inhibitor of CPG2, it was diluted with the assay buffer. The final concentration of the solution depended on the solubility of the material in the buffer. The solution was stored at -20°C . For a standard assay, assay buffer ($900\ \mu\text{L}$) is pipetted into a cuvette and $0.6\ \text{mM}$ MTX solution ($100\ \mu\text{L}$) is added. If an aliquot of a test solution is added to the cuvette, the volume of assay buffer added is reduced by the corresponding amount.

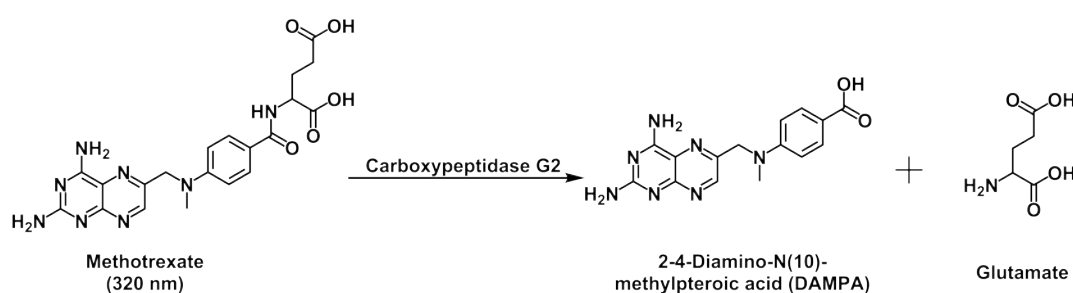


Figure 2.2: Hydrolysis of MTX measured at 320 nm.

Circular Dichroism (CD).

The protein sample was prepared in 20 mM Tris using between 0.1 and 0.2 mg/mL protein concentration. Measurements were recorded using a J-715 spectrophotometer or a Jasco J-1500 spectropolarimeter (Jasco, UK) using a 0.1 mm pathlength quartz cuvette (Starna, UK) at 25°C . Spectra were recorded between 280 and 190 nm, with a bandwidth of 2 nm, a scanning speed of 100 nm/min and 8 number of scans. For every sample, the solvent sample was recorded and subtracted from the corresponding sample spectrum.

2.2.10 NMR Experiments.

^{15}N - ^1H Heteronuclear Single Quantum Correlation (HSQC) Experiment.

It should be mentioned that CPG2₂₃₋₄₁₅, CPG2_{CAT}, CPG2_{CAT'} and CPG2_{DIM} without a his-tag were used for NMR studies (See Table 2.3). Each NMR samples of 180 μL contained approximately 0.4 mM protein dissolved in 20 mM Tris buffer (pH 7.2) containing 100 mM NaCl. Spectra were collected using 3 mm diameter NMR tubes made for the Bruker MATCHTM System on either a 700 MHz Bruker Avance spectrometer (Bruker, UK), equipped with a cryoprobe, housed at the University of Warwick or on a 900 MHz Bruker AVANCE III spectrometer (equipped with a cryoprobe) housed at the Henry Wellcome Building for NMR (HWB-NMR) at the University of Birmingham, UK. Two-dimensional ^{15}N - ^1H heteronuclear single quantum correlation (HSQC) spectra using hsqcf3gppsi (Bodenhausen and Ruben 1980; Piotto et al. 1992; Sklenar et al. 1993) and trosytf3gppsi (Czisch and Boelens 1998; Pervushin et al. 1998; Meissner et al. 1998; Weigelt 1998; Rance et al. 1999; Zhu et al. 1999) were acquired at 298 K, 303 K and 310 K with 2048 \times 128 data points and 256 scans at 700 MHz and/or 900 MHz. Water suppression was achieved using 3-9-19-pulse sequence with z-gradients (Piotto et al. 1992; Sklenar et al. 1993). The carrier positions were set to 114.994 ppm for ^{15}N and 4.719 for ^1H , the spectral widths were 14 ppm in the ^1H the dimension and 40 ppm in the ^{15}N dimension. Spectra recorded were referenced to residual water. Data were processed using TOPSPIN 3.1 by Bruker (Coventry, UK) and analysed using

NMRFAM-SPARKY (Kneller and Kuntz 1993; Lee et al. 2015).

Saturation-Difference-Transfer NMR Experiment (STD-NMR experiment).

Buffer, protein stock and ligand stock solution preparation. A 5 mL buffer solution (20 mM deuterated-Tris, 0.1 M NaCl, 0.2 mM ZnCl₂, pH7.2) in 100% D₂O was prepared. Unlabelled CPG2 protein (see Section 2.2.3) was lyophilised for STD-NMR experiments and dissolved in the appropriate amount of buffer solution to make up a 40 μ M stock solution when required and a final ligand stock solution of 5 mM was prepared (ligands were provided by Mologic Ltd, Bedford).

STD-NMR: set-up. All NMR spectroscopy experiments were performed on a Bruker Avance 700 MHz instrument equipped with a 5 mm cryoprobe. NMR samples were prepared in 180 μ L of 100% D₂O buffer containing 20 mM tris buffer pH 7.2 (uncorrected for D₂O), for the CPG2 samples.

For STD NMR experiments, a pseudo-2D version of the STD NMR sequence was used for the interleaved acquisition of off- and on-resonance spectra: 1D STD with spoil and T2 filter using excitation sculpting for water suppression (stddiffesggp.3). Off- and on-resonance spectra were acquired separately and subtracted manually to obtain the STD spectrum. For selective saturation, the on-resonance frequency was set to -301 Hz (-0.43 ppm) or -700.24 Hz (-1 ppm), whereas the off-resonance frequency was set to 30,000 Hz (42.8 ppm). Spectra recorded were referenced to residual water. In all cases, blank experiments (in the

absence of protein), were performed to test the lack of direct saturation to ligand protons and select the appropriate on-resonance frequency. Saturation times was set at 2 s and relaxation delay 3 s. Typically, the number of scans was 8, with 4 dummy scans.

Binding isotherms from STD amplification factors. The concentration of protein used was 20 μ M of CPG2 for ligand binding. Ligands were titrated into the corresponding protein samples from concentrated stock solutions. Each time a y μ L of ligand was added, also y μ L of 40 mM solution of CPG2 was added in order to keep the concentration of CPG2 constant throughout the titration. In STD-NMR experiments, at a given ligand concentration, the STD intensity of a given proton can be considered to be directly proportional to the fraction of bound ligand (f_B^L) (Angulo et al. 2010; Viegas et al. 2011).

$$f_B^L = \frac{[PL]}{[L]} = \frac{(I_o - I_{\text{sat}})}{I_o} \quad (2.3)$$

STD-AF, defined by Mayer and Meyer as the product of f_B^L by the ligand excess ($EL = \frac{[L]}{[P]}$). Hence:

$$STD - AF = \frac{(I_o - I_{\text{sat}})}{I_o} \times \frac{[L]}{[P]} \quad (2.4)$$

Plotting STD-AF values at increasing values of [L] gives rise to a Michaelis-Menten hyperbolic dose-response as described by equation:

$$STD - AF_{[L]} = \frac{STD - AF_{\text{max}}[L]}{[L] + K_D} \quad (2.5)$$

Where $STD - AF_{max}$ is a dimensionless scaling factor representing the maximum STD amplification factor for the monitored signals. Mathematical fitting yields to both parameters: K_D and $STD - AF_{max}$.

For each ligand concentration, STD-AF build-up curves were recorded. To determine K_D values following the standard procedure, the STD-AF values (at a given saturation time) of a ligand proton were plotted as a function of the concentration of ligand, and the resulting curve was mathematically fitted to the dose-response equation (2.5), to yield to K_D .

Equilibrium Kinetics of Binding



Fraction of bound ligand is $f_B^L = \frac{[PL]}{[P] + [L]}$, but since $[L] \geq [P]$, we can make the following approximation:

$$f_B^L \approx \frac{[PL]}{[L]} \quad (2.7)$$

2.2.11 X-ray Crystallography.

Growth of Protein Crystals.

Precipitants were used to grow proteins crystals by precipitation from aqueous solutions. There are diverse precipitants, including salts, organic solvents and polymers (e.g. PEG) and combinations of various types of precipitants can be used. Other criterion also influence crystal growth, such as precipitant

concentration, pH and temperature (Boistelle and Astier 1988). In order to find appropriate conditions for crystallisation of a protein, screening was achieved over a range of different conditions to see when crystals were formed and after that optimisation of those conditions was usually needed to improve the crystals.

The procedure of growing crystals usually involves addition of the precipitants to a aqueous solution of the protein in a concentration just below what is needed for precipitation. After that, the water is evaporated slowly until precipitating conditions was reached. The method used for crystallisation was vapour diffusion. This method is based on the technique described by Unge (1999). A droplet of purified protein solution mixed with a crystallisation solution is deposited onto a cover glass. The crystallisation solution often consists of buffer, salt and precipitant. The cover is suspended above a reservoir of crystallisation solution. Due to concentration gradient, the water from the drop evaporates towards the reservoir, thus lowering the volume of the drop and increasing the protein concentration. This will result in supersaturation of the protein solution and crystals will start to form (Unge 1999). Vapour diffusion can be performed in different ways, for example hanging drop and sitting drop, as shown in Figure 2.3.

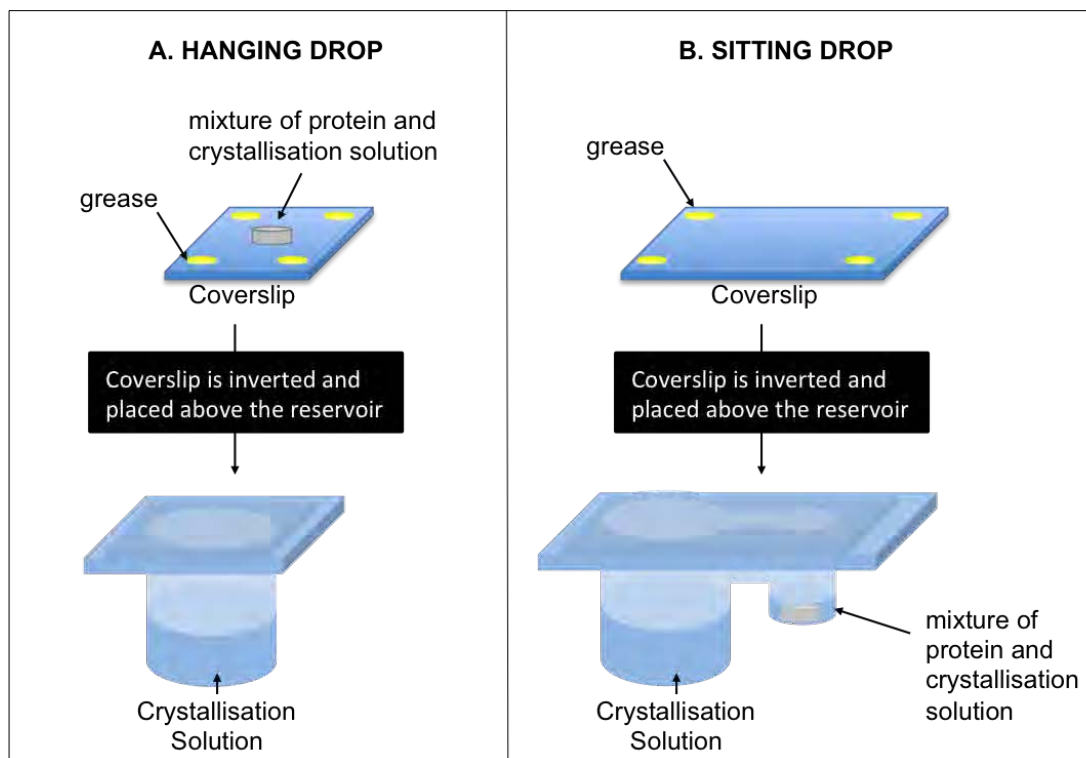


Figure 2.3: Different vapour diffusion system. A) Hanging drop and B) sitting drop setups. Adapted from Unge (1999).

Co-Crystallisation.

Protein-ligand interactions can also be studied with X-ray crystallography. For this, crystals of protein-ligand complexes are needed. There are different ways of achieving this, including soaking and co-crystallisation. These two methods were used (Hassell et al. 2007).

Soaking: This method implicates transferring the protein crystals into crystallisation solution that contains the ligand. The ligand can diffuse through water channels in the crystal and thereby reach the active site. A requirement for this method is that the protein crystals are functional for binding ligands. This can vary for different ligands and the soaking process can be affected by solubility, size and shape of the ligand.

Co-crystallisation: The ligand and protein are mixed and crystallised together. This method is more time-consuming and requires more protein. The crystal structure can differ with different ligands and might not be the same as for crystals of the protein without ligand, and requires a new screening to achieve crystals.

Crystal Analysis.

Hanging drop setups were prepared manually using 24-well tissue plates as described in Chapter 3. However, sitting drop trays were set up using Honeybee® 963 and Mosquito® LCP³ crystallisation robots to screen 96×1 mL trays. For each ligand (CP06 and CP67) three 96-condition crystallisation screens were used; JCSG+, PACT and Morpheus purchased from Molecular Dimensions Ltd. (Suffolk, UK). Crystallisation screening was performed using reservoir volumes of 50 μ L, and 2 μ L sitting drops comprising 1 μ L of protein solution and 1 μ L of reservoir solution. The six trays were stored in a constant temperature room at 18°C and inspected daily using an Olympus SZ4045TR-PT microscope.

In both setups, generated crystals were tested with X-rays⁴. Crystals that diffracted to better than 4 Å resolution were taken to the the Diamond Synchrotron Source in Oxford. Model building for the crystals, that were tested so far at the synchrotron, was done by molecular replacement, as we already

³Dr Dean Rea (School of Life Sciences, University of Warwick) set up sitting drop screening using Mosquito LCP crystallisation robot.

⁴Dr Karen Ruane (School of Life Sciences, University of Warwick) tested crystals for diffraction, collected the data at Diamond Synchrotron Source in Oxford and solved the structure.

had a coordinate set to work with (PDB: 1CG2) (Rowse et al. 1997).

Chapter 3

Protein Preparation for NMR

Spectroscopy and X-ray

Crystallography Studies

3.1 Introduction.

Due to its wide use in cancer therapy, there has been a massive need for high yield expression of CPG2. The first goal of this work was development of an *E. coli*-based protocol for overexpression of CPG2 in a soluble form. *E. coli* was a more attractive alternative to *Pseudomonas* sp. RS-16, previously used to produce CPG2 for crystallography studies, as it offers shorter culturing time, fast high density cultivation and easier genetic manipulation than *Pseudomonas*, the natural CPG2 source (Rowsell et al. 1997). Because *Pseudomonas* and *E. coli* differ greatly in their relative codon-usage, the CPG2 gene from *Pseudomonas* sp.

strain RS-16 was codon-optimised and cloned into a pET28a vector for maximum expression in BL21 (DE3) *E. coli* strain (Goda et al. 2009). In spite of all these qualities, expression of CPG2 with *E. coli* as the host resulted in high yields of the protein, but in its insoluble form (Goda et al. 2009). Such proteins, forming dense aggregates (i.e. inclusion bodies), are frequently improperly folded and thus biologically inactive (Singh and Panda 2005).

In the context of this project, it was vital to express high-yields of CPG2 in its soluble and active form in order to preserve the integrity of the protein fold for future structural studies.

3.2 Expression of Codon-Optimised Full Length CPG2 in *E. coli* Leads to Intermittent Activity and Existence of Two CPG2 Subspecies.

A number of tactics for the redirection of proteins from inclusion bodies to soluble cytoplasmic fractions are described in the literature. In general, they can be divided into methods where the expression strategy is modified to obtain soluble expression and methods where the protein is refolded from inclusion bodies. In our study, refolding from inclusion bodies, as described previously by Goto and Kay (2000), was considered undesirable due to lengthy and complex refolding protocols and the likelihood of poor recovery yields, increasing the expense of

each sample upon expression in isotopically $^{13}\text{C}/^{15}\text{N}$ -enriched minimal media. Accumulation of denatured protein can be minimised by optimisation of a number of environmental factors (Sørensen and Mortensen 2005). To increase the levels of soluble full-length CPG2 *in vivo*, a range of diverse expression conditions were initially tested using small-scale cultures.

3.2.1 Testing Different *E. coli* Strains for Higher Solubility.

Firstly, two *E. coli* mutant strains of BL21 (DE3), namely the C41 (DE3) and C43 (DE3) strains known to promote soluble protein expression in previous studies, were tested at 20°C using 1 mM IPTG to induce protein expression (Figure 3.1) (Miroux and Walker 1996). Interestingly, C43 (DE3) cell growth was limited: the optical density never exceeded OD₆₀₀ of 0.1 and never entered the exponential growth phase. However, C43 (DE3) expressed CPG2 (Figure 3.1-A). No enhancement of soluble CPG2 expression was observed and overall expression in C43 (DE3) was comparable to that observed in BL21 (DE3), hence BL21 (DE3) was selected as the host (Figure 3.1-B).

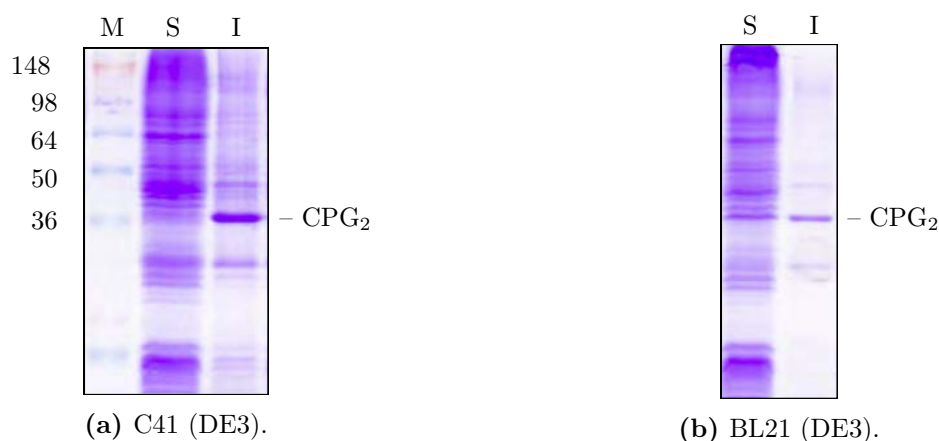


Figure 3.1: Expression in different *E. coli* strains. SDS-PAGE gel showing (*M*) protein marker, (*I*) insoluble and (*S*) soluble fractions after expression of CPG2 using A) C41 (DE3) and B) BL21 (DE3) with 1 mM IPTG at 20°C for 4 hours.

3.2.2 Effect of Changes in Temperature, IPTG Concentration and Induction Duration.

Reduction of the expression temperature and IPTG concentration are also commonly used approaches to limit *in vivo* protein aggregation in BL21 (DE3) (Weickert et al. 1996). However, a sudden decrease in temperature can negatively affect, transcription and replication (Shaw and Ingraham 1967). Hence, it is vital to screen for different induction concentrations and expression durations that would result in higher levels of protein expression at this temperature.

Thus, a number of IPTG induction concentrations ranging from 0.1 mM to 0.9 mM have been tested as seen in Figure 3.2. Induction for 4 hours expression (Figure 3.2-A) indicated lower expression of the protein compared to 24 hours (Figure 3.2-B) and 48 hours expression (Figure 3.2-C). In general, IPTG concentration had little impact on CPG2 expression level, but the length of expression duration had a significant impact on the expression level. At

concentrations above 0.7 mM at the different expression durations, we clearly observed a lower expression of soluble protein, but the protein band in the insoluble fraction had the same intensity when induced at concentration as low as 0.5 mM. Hence the yield, when using BL21 (DE3) as a host, was significantly improved by the use of low induction temperature (15°C) and a 24 hours expression duration in both rich and minimal media (Figure 3.2-B).

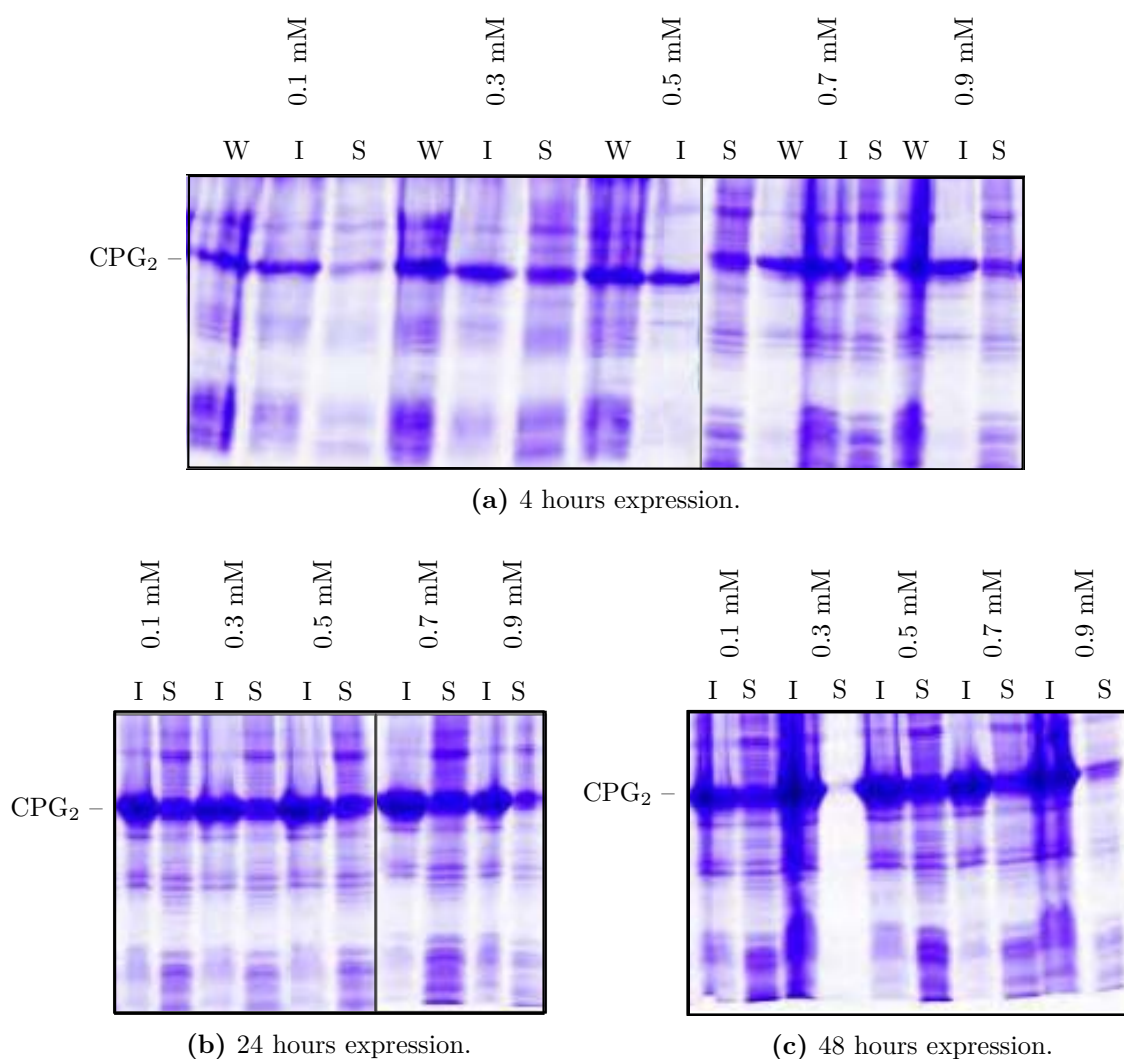


Figure 3.2: Testing different IPTG induction concentration and expression duration at 15°C. SDS-PAGE gel showing (W) whole cells, (I) insoluble and (S) soluble fractions showing expression levels of CPG₂ after A) 4, B) 24 and C) 48 hours incubation periods with 0.1, 0.2, 0.5, 0.7 and 0.9 mM IPTG concentrations. SDS-PAGE samples were normalised to OD₆₀₀ of 6.

3.2.3 Presence of Possibly Two CPG2 Isoforms.

Full-length CPG2 was successfully expressed and purified from BL21 (DE3) cells under control of the T7 lac promoter system, grown from either 2×YT media for natural isotopic abundance. The ideal assay to evaluate the quality of the soluble CPG2 produced was the enzymatic activity based on cleavage of the substrate MTX (see Section 2.2.9 of Chapter 2).

However, the enzymatic activity of replicate batches of CPG2 preparations showed high inconsistency (Figure 3.3-A). SDS-PAGE gels indicated the presence of two bands close to the expected molecular weight of CPG2 (see bands indicated by arrows in Figure 3.3-C, left panel). In-gel tryptic digestion and protein identification using mass spectrometry was performed to identify the band of interest, as it was expected that the two populations corresponded to CPG2 with and without signal peptide. Results showed that both bands corresponded to different isoforms of full-length CPG2 (Figure 3.3-C, right panel), although the intensities of the bands were quite different. It was thought that perhaps only one of the proteins was contributing to the enzymatic activity, which would account for the poor batch-to-batch reproducibility of the activity assay (Figure 3.3-A). CPG2 contains an N-terminal hydrophobic signal sequence that directs it to the periplasm of *Pseudomonas* sp. Strain RS16. The protein identification using mass spectrometry data¹ (Figure 3.3-C, right panel) suggest that reducing the temperature of induction (and thus expression level) did not lead to expression of CPG2 in its soluble, periplasmic form in the

¹Dr Suzanne Slade (School of Life Sciences, University of Warwick) performed tryptic digest, protein identification and analysis by Nano LC-ESI-MS/MS.

mature form containing no signal sequence.

Estimation of the secondary structure content via DichroWeb analysis of the CPG2 CD spectrum (Figure 3.3-B, see dotted line) yielded 9% α -helical and 47% β -sheet content (Whitmore and Wallace 2004; Whitmore and Wallace 2008). Comparison of these to the values obtained from the crystal structure (PDB ID: 1CG2; 38% helix, 26% sheet) suggest that the full-length CPG2 is improperly folded (Rowse et al. 1997).

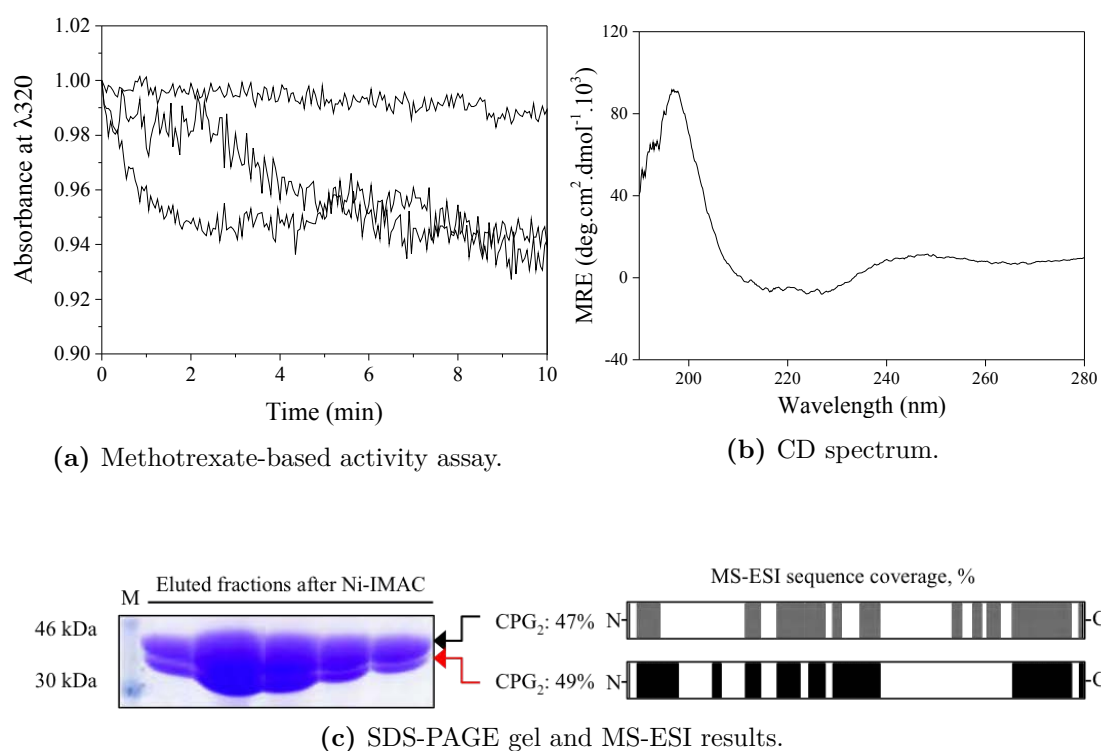


Figure 3.3: Expression of CPG2 in the pet28a vector revealing the presence of two immature isoforms. A) Activity assays of the purified CPG2 using the same construct indicating inactivity over time, B) CD spectrum of immature CPG2 and C) SDS-PAGE gel (left panel) illustrating the presence of two CPG2 isoforms during purification. Mass spectrometry (right panel) analysis of the two bands of interest (results for bands 1 and 2 are shown respectively in top and bottom strips) with 47% and 49% similarity to immature (full-length) CPG2 protein sequence. It should be noted that the signal peptide is present in both bands.

3.3 Production of Soluble and Active CPG2.

Two approaches were then considered in order to express biologically active soluble protein using BL21 (DE3) as a host. The first was removal of the N-terminal signal peptide sequence, which is unlikely to be recognised by *E. coli*, to generate soluble cytoplasmic protein. The second was replacement of the native signal peptide (residues 1-22) with a leader sequence (i.e. pelB) used for periplasmic localisation in *E. coli* bacterial strains (Choi and Lee 2004). The latter usually produces much lower yields since not all expressed protein is secreted into the periplasm.

Therefore, the first option was prioritised, as it restricts the localisation of the protein to one compartment of the cell (cytoplasm).

3.3.1 Removal of the Signal Peptide Induces Expression of a Single, Soluble CPG2 Species.

The pET28a plasmid containing the full-length CPG2 gene was used as a template to generate a cDNA sequence of CPG2 without the leader sequence (CPG2₂₃₋₄₁₅), which was then subcloned into the pET151-D-TOPO vector (Invitrogen) and expressed and purified as detailed in Materials and Methods (Figure 3.4).

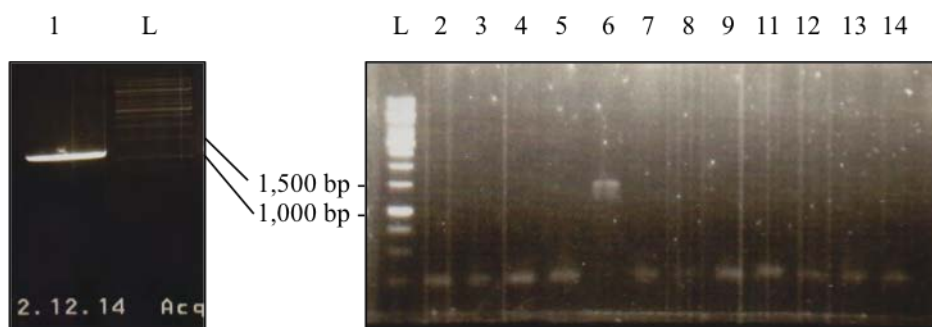


Figure 3.4: Agarose gel electrophoresis (2%) of PCR products. The left panel shows amplification of the gene of interest (without signal peptide, CPG2₂₃₋₄₁₅) using pet28a vector (expected size fragment at 1.3 kb). *lane 1*: PCR product (expected size 1.3 kb) and *lane L*: 1kb DNA ladder (Generuler). Right panel displays colony screening by PCR for gene of interest using T7 forward and reverse primers. Only colony 6 displayed a positive clone (expected size fragment at 1.4 kb).

Removal of the signal peptide led to a single-band (Figure 3.5-A, lanes 1-5) and a 6-fold increase in protein expression yield (Table 3.1).

Table 3.1: Difference in expression yield between wild-type CPG2 and CPG2₂₃₋₄₁₅.

Constructs	His-tag Terminus	Activity U/mL	Specific Activity U/mg	Protein Yield [†]		
				Unlabeled (2×YT)	¹⁵ N (M9)	² H/ ¹³ C/ ¹⁵ N (M9)
(His) ₆ -CPG2	N	32.9	5.6	38.6	nd [‡]	nd
(His) ₆ -CPG2 ₂₃₋₄₁₅	N	100.26	75.5	251.6	nd	70-109.42

[†] Yields of purified proteins are in mg/L of culture.

[‡] Not determined.

His₆-CPG2 was purified and the his-tag was cleaved according to the protocol described in Section 2.2.6. Because Ni²⁺ IMAC purification was not effective at isolating CPG2 from His₆-CPG2, an extra gel filtration purification step was introduced (Figure 3.5-B). The purified CPG2₂₃₋₄₁₅ was then tested for its ability to catalyse the hydrolysis of MTX. Figure 3.5-C demonstrates that

CPG2₂₃₋₄₁₅ was able to hydrolyse MTX with an enzymatic activity of 100.26 U/mL (Table 3.1), indicating that the recombinant enzyme produced in its soluble form is biologically active. Circular dichroism measurements of CPG2₂₃₋₄₁₅ in the wavelength range between 190 nm and 280 nm were performed, and suggest 53% α -helical / 20% β -sheet content after fitting of the data using DichroWeb (Figure 3.5-D) (Whitmore and Wallace 2004; Whitmore and Wallace 2008). These values are in better agreement with those predicted from the crystal structure (i.e. 38% helix, 26% sheet), and suggest that the presence of the signal peptide was preventing proper folding of the enzyme (Rowsell et al. 1997).

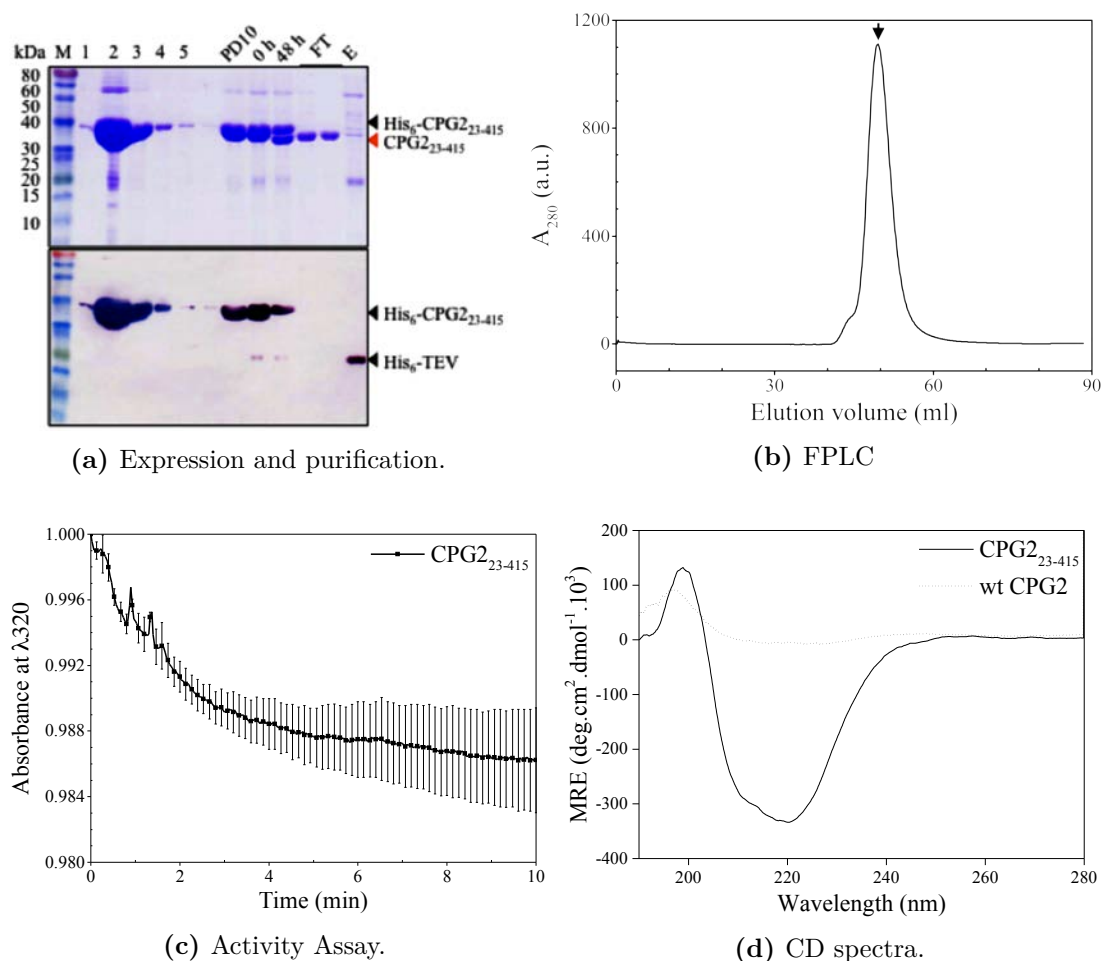


Figure 3.5: Expression and characterisation of CPG2₂₃₋₄₁₅. A) Top: SDS-PAGE gel purification, bottom: western blot, *M* denotes the protein marker, 1 to 5 are the eluted sample after nickel chromatography affinity purification, PD10 indicates presence of the protein after desalting, lane *0 h* refers to incubation with His-TEV protein at 0h and the following lane after *48 h* of incubation, lanes *FT* and *E* show the his-tag free CPG2 and eluted His-TEV collected in the flow-through and the eluent respectively after a second round of nickel-affinity chromatography column. B) Gel filtration chromatography the arrow indicates elution of untagged protein. C) CPG2₂₃₋₄₁₅ enzymatic activity assay against methotrexate substrate. D) Circular dichroism in the 190-280 nm region for the wild-type CPG2 and CPG2₂₃₋₄₁₅.

3.3.2 Purification of CPG2₂₃₋₄₁₅ by Gel Filtration Chromatography.

Elution fractions from IMAC containing CPG2₂₃₋₄₁₅ were next applied to PD10 desalting columns (see Section 2.2.6) and further purified by Gel Filtration Chromatography (see Section 2.2.7). Improved separation of CPG2₂₃₋₄₁₅

proteins was achieved by applying samples that had been concentrated using 30 kDa MWCO concentrators (see Section 2.6.8) to a HiLoad 16/600 Superdex 75 g GL column (separation range 1-120 kDa), under an isocratic gradient of 100% A (20 mM Tris, 100 mM NaCl, pH 8). As shown in Figure 3.6, the protein eluted as a double peak at elution volumes of 62.9.4 mL for dimeric CPG2₂₃₋₄₁₅ and 52.4 mL for monomeric CPG2₂₃₋₄₁₅.

By applying Gel Filtration standards to the column and running at the same flow rate and with the same buffer (see Section 2.2.7), it was possible to create a calibration curve of molecular weight against elution volume, assuming a void volume of 25.0 mL (Figure 2.1). Plotting the relative retention times of CPG2₂₃₋₄₁₅ proteins on this curve gave predicted molecular weights of around 43 kDa and 103 kDa (calculated monomeric peak for CPG2₂₃₋₄₁₅ is 42,284 Da), and therefore CPG2₂₃₋₄₁₅ is expected to migrate and exist in monomeric and dimeric forms (Figure 3.6). Only the eluted fractions corresponding to the monomeric CPG2₂₃₋₄₁₅ peak were collected for NMR spectroscopy.

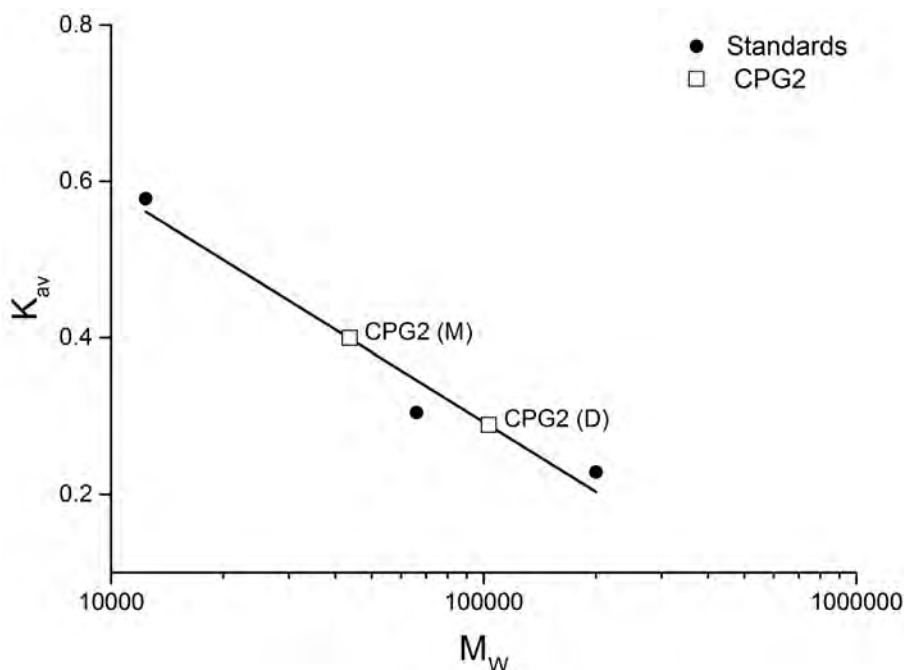


Figure 3.6: Calibration curve obtained with Gel Filtration standards on HiLoad 16/600 Superdex 75 g GL column. An isocratic gradient of 100% of buffer A (20 mM Tris, 100 mM NaCl, pH 8) was used. The calibration curve graph represents the gel phase distribution coefficient K_{av} versus the molecular weight (M_w). $K_{av} = (V_e - V_o)/(V_c - V_o)$, where V_e = elution volume, V_o = column void volume, V_c = geometric column volume. Straight line is the calibration curve calculated from the data for molecular weight standards (see Section 2.2.7 and Figure 2.1). Using the equation for the line, values for the molecular weights of CPG2₂₃₋₄₁₅ were calculated and plotted on the graph.

3.4 High-Yield Expression of Isotopically Labelled Protein for NMR Studies.

$^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -enriched protein samples are essential for high-resolution structural studies of large proteins using NMR. To achieve this, protein expression must be carried out in M9 media so the investigator has full control over the sources of carbon, nitrogen and hydrogen utilised by the cell. Expression trials were carried out for CPG2₂₃₋₄₁₅ in M9 media, followed by purification. SDS-PAGE analyses and estimation of concentration using the BCA assay indicated over 50% reduction in soluble protein yield of CPG2₂₃₋₄₁₅ upon switching the medium from 2×YT to M9 (Figure 3.7)

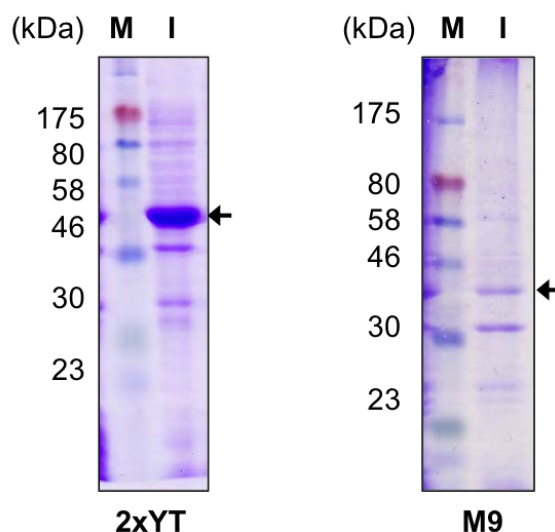


Figure 3.7: Comparison of CPG2 expression in 2×YT and M9 media. SDS-PAGE gel showing (*M*) protein marker, (*I*) induced CPG2 expression with 0.5 mM IPTG for 18 hours at 15°C in *left panel* 2×YT medium and *right panel* M9 medium. The CPG2 protein band is indicated with a black arrow.

The effect of incubation temperature on BL21 (DE3) cell growth and CPG2₂₃₋₄₁₅ expression was therefore investigated by analysing the *E. coli* optical density after inoculation, as it can be graphically translated into growth curves. Cell growth in rich medium did not reach the stationary phase and stayed in the beneficial exponential phase for expression at optical densities as high as 6 (Figure 3.8, solid lines). Conversely, when cells were grown in M9 (Figure 3.8, dotted lines) stationary phase was reached more rapidly. 2×YT medium is abundant in critical nutrients, necessary for the bacteria to multiply quickly without depleting the host cells of these rudimentary nutrients. As opposed to 2×YT, M9 is a minimal medium only containing salts and nitrogen, which can usually be supplemented with labelled carbon and / or amino acids to make a complete medium for labelled NMR sample preparation. Therefore, cell growth in M9 correlates to a quicker formation of an unfavourable environment for

bacteria, due to depletion of lower level of nutrients to start with, which may explain the lower protein yield. The death phase was not reached in the first 15 hours observed, in either of the media used (Figure 3.8). The latter reveals that if cells are taken at the stationary phase and moved to a fresher M9 media, they still have a possibility of reaching the exponential phase.

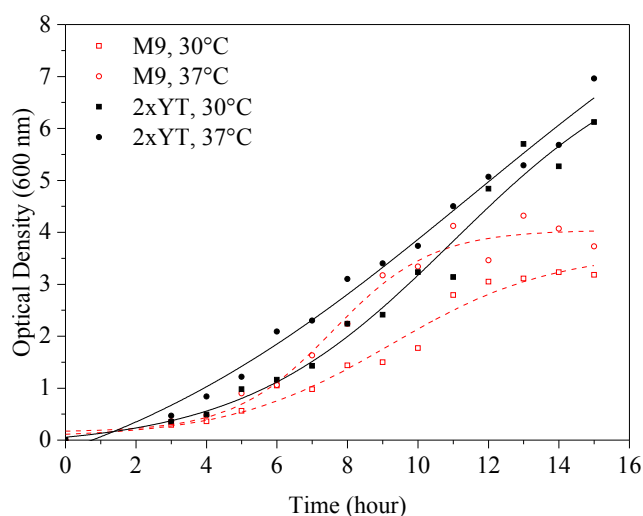
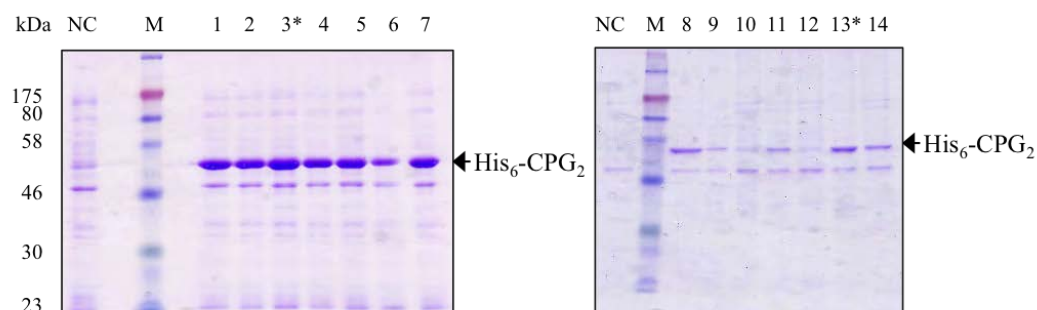


Figure 3.8: Growth curves of cultures post-inoculation with minimal (M9) and rich 2×YT media at different temperatures. Cell growth was tested at different incubation temperatures, 30°C and 37°C, with the OD₆₀₀ of bacterial growth samples being recorded at one hour intervals. Bacterial cells post-inoculation were able to grow at temperatures as low as 30°C and as high as 37°C in both media, with 37°C being optimal. 2×YT growth curves obtained are consistent with 2×YT being a nutritionally rich media. On the other hand, as expected, cell growth in M9 rapidly attained the stationary phase at an OD₆₀₀ of 3.5.

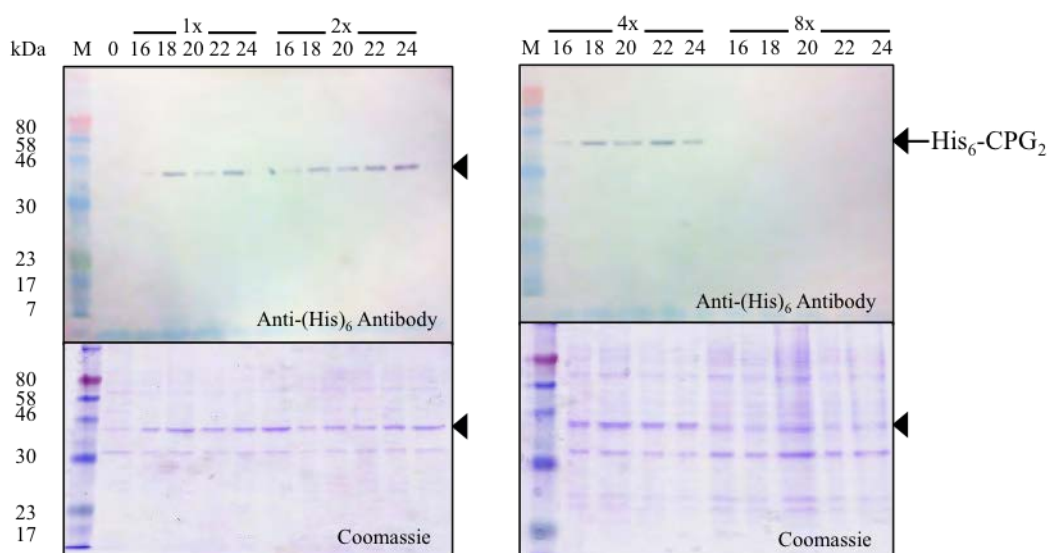
To reduce the cost of each labelled sample, the results arising from the effect of M9 on cell growth (Figure 3.8) and the innovative expression method described by Sivashanmugam et al. (2009) were used to increase cell density without interfering with the expression vector (referred to herein as the cell condensation method). Firstly, a double colony selection was performed (Figure 3.9-A) (Marley et al. 2001). Rich medium (50% D₂O) was inoculated with freshly transformed cells, grown at 37°C overnight, and spread onto an M9 agar plate containing 50% D₂O. Each colony was checked for expression in 10 mL M9

with 50% D₂O through the usual IPTG-induction method (Figure 3.9-A, left panel). The colony exhibiting the highest-level CPG2₂₃₋₄₁₅ expression was used for a second round of selection (Figure 3.9-A, left panel, colony 3). Since high-levels of perdeuterated protein were required, the second round of selection was conducted in 100% D₂O. The colony displaying the highest levels of CPG2₂₃₋₄₁₅ expression under these conditions was picked (Figure 3.9-A, right panel, colony 13). The colony-to-colony variability in CPG2₂₃₋₄₁₅ expression levels highlights the need to screen colonies at this stage to obtain optimal expression levels.

After double-colony selection, the cell condensation method was used (Figure 3.9-B) (Sivashanmugam et al. 2009). The cells were grown to an OD₆₀₀=1 in the absence of isotopes in rich medium in order to generate a large amount of cell mass, then gently spun down and resuspended in a condensed volume of isotope-enriched minimal media for expression. The degree of condensation yielding maximal expression differs depending on the protein expressed, and must be investigated. Small-scale time-course experiments were carried out for 1-fold, 2-fold, 4-fold and 8-fold condensations (respectively 100%, 50%, 25% and 12.5% of the initial volume), and CPG2₂₃₋₄₁₅ expression was evaluated using SDS-PAGE and western blots (Figure 3.9-B). Visual inspection of blots indicated that 2-fold and 4-fold condensation produced optimal yield and resulted in to a four-fold decrease in the cost of protein samples (as compared to conventional labelling, e.g. 1-fold) and generated 70-109.42 mg of labelled protein per litre of bacterial culture (Table 3.1).



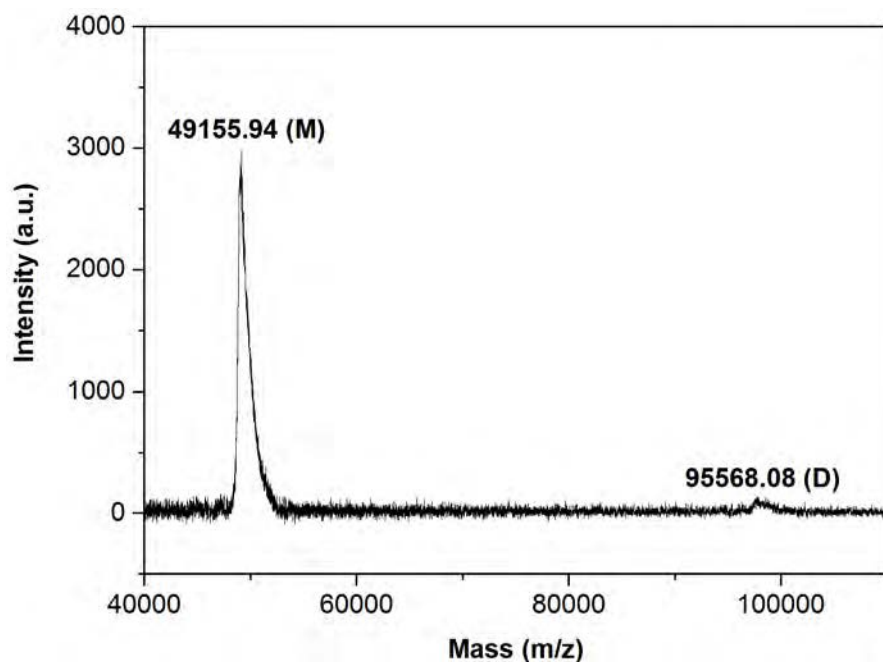
(a) Double Colony Selection.



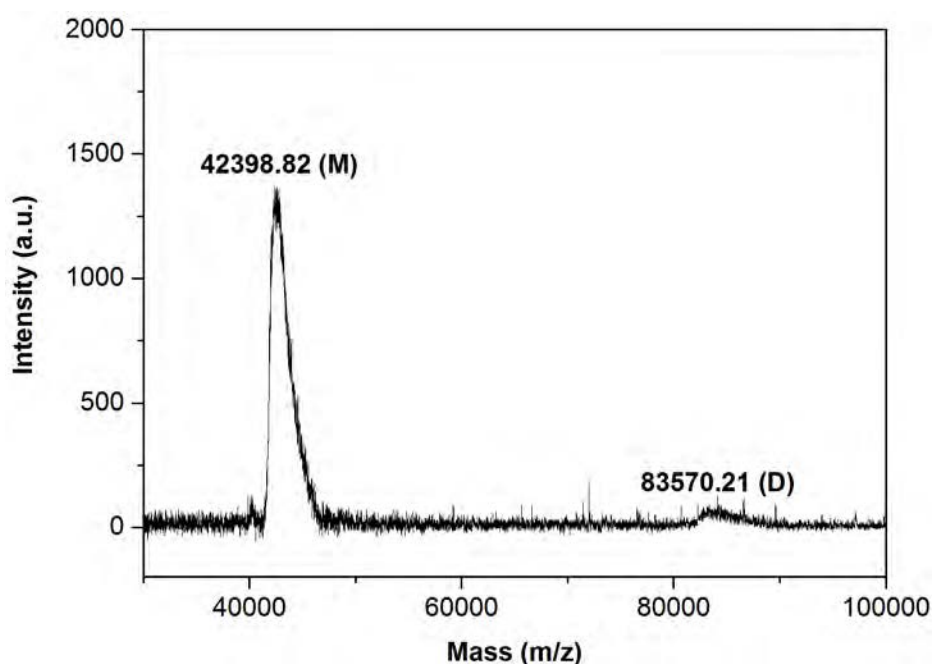
(b) Condensation method.

Figure 3.9: SDS-PAGE gels of double colony selection and condensation method. To evaluate the amount of CPG2₂₃₋₄₁₅ on the western blots, the volumes of samples loaded onto the gels were normalised with respect to the concentration of the cells in the pre-induction sample (lane NC and 0). A) SDS-PAGE of protein expression of CPG2₂₃₋₄₁₅ after first colony (lanes 1-7) and second colony selections (lanes 8-14). The black arrow indicates CPG2₂₃₋₄₁₅ expression levels during the double selection method in the different colonies and each asterisk indicated the colony selected after each step. B) Effect of condensation on CPG2₂₃₋₄₁₅ expression level. 1x, 2x, 4x and 8x labelled lanes indicate the condensation factor which corresponds to 100 mL, 50 mL, 25 mL, and 12.5 mL relative to the initial volume of rich medium used for cell growth (100 mL) pre-induction (lane 0), post-induction (lanes 16, 18, 20, 22 and 24, IPTG induction duration hours). These concentrations were investigated for CPG2₂₃₋₄₁₅ protein expression (indicated by an arrow) using Western-Blots (top) and SDS-PAGE analysis (bottom).

Isotopically labelled CPG2₂₃₋₄₁₅ migrated similarly to unlabelled CPG2₂₃₋₄₁₅ on SDS-PAGE gels, thus MALDI-TOF mass spectrometer was used to examine whether *in vivo* incorporation of heavy ²H, ¹³C and ¹⁵N labels into CPG2₂₃₋₄₁₅ was acceptable for NMR studies. Figure 3.10-A shows the result of the mass spectrometer analysis of pure CPG2₂₃₋₄₁₅. Assuming 100% label incorporation, the theoretical molecular weight is 49,193 Da, a 6,809 Da greater than the theoretical unlabelled protein molecular weight of 42,384 Da (3.10-B). The main peak observed in Figure 3.10-A was at 49,155.94 Da illustrates the label enrichment to be >90%, which is suitable for NMR analyses. The absence of 100% labelling is possibly due to the labile protons in the amide groups of the protein. The broadness of the peak is primarily due to acquisition of the mass spectrometer data in linear mode, which is used to keep the protein intact (impossible using reflector mode). In addition, it also demonstrates that there was a broad distribution of labelled CPG2₂₃₋₄₁₅ with varying degrees of label incorporation, as expected for protein expressed in labelled media. Mass spectrometer also showed the presence of higher oligomeric state that required a gel filtration step (Figure 3.5-B).



(a) [^2H , ^{13}C and ^{15}N]-enriched CPG₂₂₃₋₄₁₅ preparation.



(b) Unlabelled CPG₂₂₃₋₄₁₅ preparation.

Figure 3.10: Maldi-TOF mass spectrometer *A*) labelled and *B*) unlabelled CPG₂₂₃₋₄₁₅. *A*) Main peak observed at 49155 Da is in close agreement with expected molecular weight of 48193 Da. *B*) Main peak observed at 42570 Da is in close agreement with theoretical molecular weight of 42384 Da. In both cases, the presence of dimer (D) is noticed.

3.5 Stability of CPG2₂₃₋₄₁₅ is Dependent Upon Buffer Conditions.

While we had successfully expressed CPG2₂₃₋₄₁₅ in a soluble form in *E. coli*, the resulting protein displayed limited long-term stability (activity deteriorated 24 hours post-purification) and a propensity to aggregate over time in Tris buffer containing 50 mM NaCl (pH 7.2). To create a stable CPG2₂₃₋₄₁₅ sample for NMR analyses (which must remain in its active, soluble state for days / weeks), a variety of buffer conditions were tested. Salt concentrations were varied from 50 mM to 500 mM to prevent loss of sensitivity due to increased sample conductivity when using the cryogenic NMR probe (Kelly et al. 2002). This complicated matters, as salt concentrations can significantly affect protein solubility and aggregation. We were also restricted from use of glycerol, commonly used to increase protein stability, in the buffer as it can further reduce sensitivity in downstream NMR measurements due to a reduction in the tumbling time of the protein.

Three buffers were tested, which were similar apart from the NaCl concentration Table 3.2.

Table 3.2: Buffer composition for buffers 1-3. Buffers 1-3 used to investigate the effect of NaCl concentration on activity and fold of CPG2₂₃₋₄₁₅.

	[NaPi], mM	[KH ₂ PO ₄], mM	[KCl], mM	[NaCl], mM	pH
Buffer 1	10	2	2.7	500	7.2
Buffer 2 (Gomori)	20	-	-	50	7.3
Buffer 3 (1×PBS)	10	1.8	2.7	137	7.3

As shown in Figure 3.11, the activity of CPG2₂₃₋₄₁₅ was significantly affected by changes in salt concentration (Table 3.2). The enzyme was not active in buffers 1 and 2 (500 mM and 50 mM NaCl, respectively), but did show wild-type levels of activity in buffer 3 (137 mM NaCl) and this activity was maintained over days. These changes in activity were not due to changes in global enzyme structure, since CD spectra measured in all three buffers were largely identical (Figure 3.11-B) and resulted in very similar secondary structure content as estimated using DichroWeb (Figure 3.11-C) (Whitmore and Wallace 2004; Whitmore and Wallace 2008).

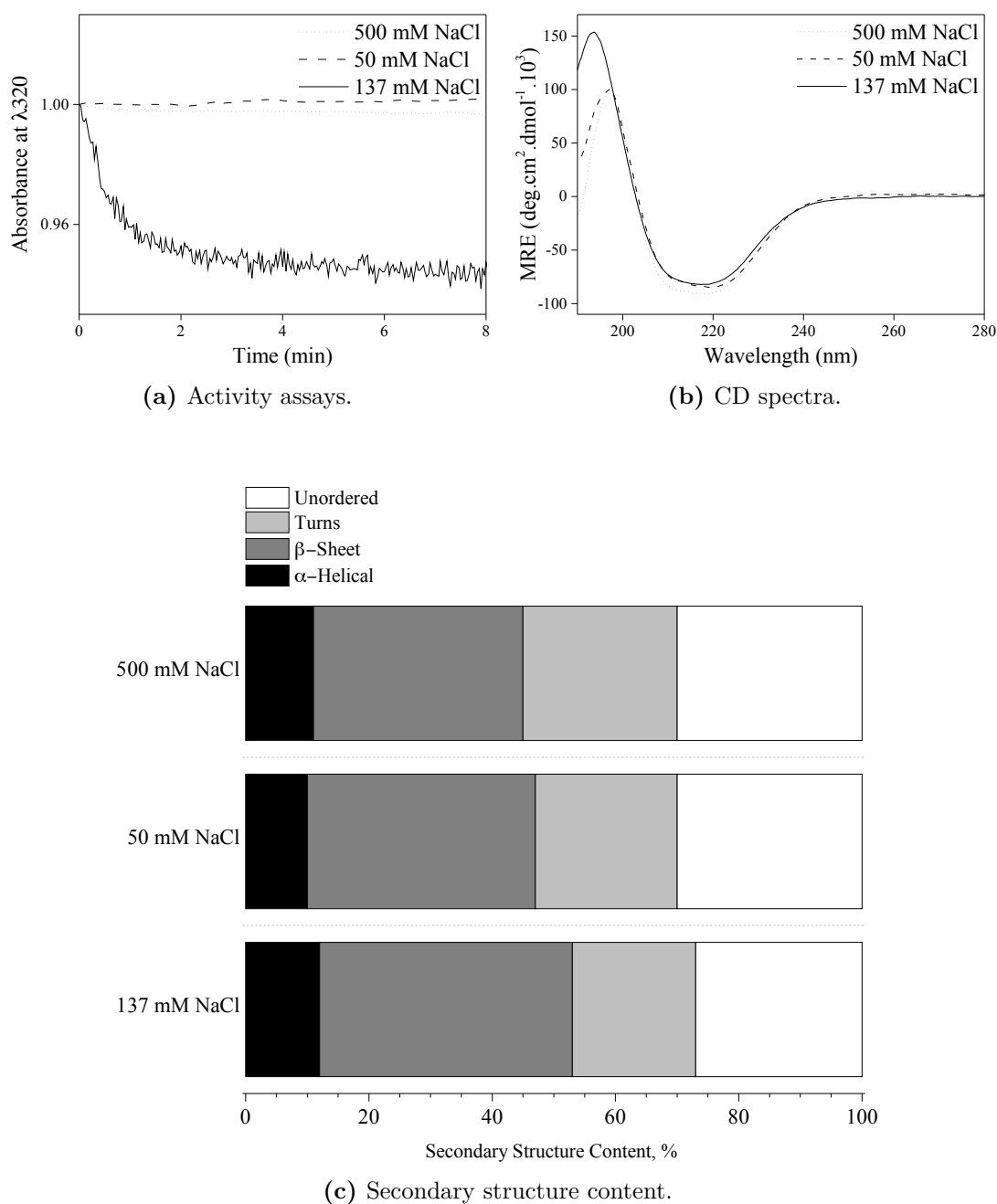


Figure 3.11: Characterisation of CPG2₂₃₋₄₁₅ in buffers 1-3, primarily differing in their NaCl content. A) Activity assay of CPG2₂₃₋₄₁₅ in Buffers 1-3 (see Table 3.2 for buffer composition) against its substrate methotrexate. B) Circular dichroism spectra. C) Percentage of secondary structure of CPG2₂₃₋₄₁₅ in the three different purification buffers using DichroWeb (Whitmore and Wallace 2004; Whitmore and Wallace 2008).

To investigate whether protein solubility could account for differences in the activity, CPG2₂₃₋₄₁₅ was solubilised using a low-conductance solution of 20 mM Tris buffer (pH 7.3) containing 50 mM L-arginine and 50 mM L-glutamate in place of NaCl (Dötsch and Withers 2005). This approach has been used in the past to prevent protein aggregation, achieve higher soluble protein concentration and increase protein solubility. These buffer conditions were initially used to screen for the best NMR conditions. Figure 3.12 shows that CPG2₂₃₋₄₁₅ was active and folded 20 mM Tris buffer (pH 7.3) containing 50 mM L-arginine and 50 mM L-glutamate. Unfortunately, this buffer did not yield CPG2₂₃₋₄₁₅ stability and solubility at NMR-amenable concentrations (approximately 0.5 mM) for over 2 days in the magnet at either 25°C or 37°C, which makes it intractable for collection of standard triple resonance experiments required for sequential assignment of large proteins.

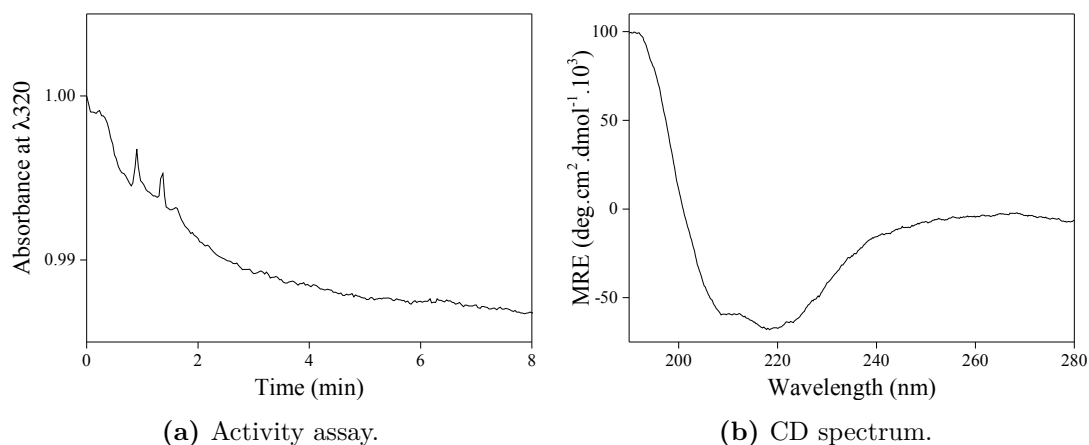


Figure 3.12: Characterisation of CPG2₂₃₋₄₁₅ in 50 mM Arg / 50 mM Glu. A) Activity assay of CPG2₂₃₋₄₁₅ in 50 mM Arg / 50 mM Glu against its substrate methotrexate. B) CD spectrum.

The buffer composition was therefore examined more comprehensively using the approach described by Ducat *et al.*, in which a typical screening technique used in crystallography studies was applied to determination of optimal sample conditions for structural NMR studies (Ducat et al. 2006). We tested 363 buffer conditions using their method, in which 1 μ l of approximately 23 mg/mL protein sample was combined with 1 μ L of various buffers. The precipitation pattern observed for CPG2₂₃₋₄₁₅ in crystallisation plates is described in Table 3.3. In general, precipitation took place at values below pH 6, except when the concentration of NaCl was at or above 200 mM. Above pH 6, CPG2₂₃₋₄₁₅ remained soluble in all buffers for at least 5 days at 18°C, except when arginine and glutamate were included in the buffer. The pH range within which no precipitation was observed over longer timescales was quite limited (essentially around physiological pH) and greatly depended on the type and concentration of salt. The analysis of the precipitation pattern clearly indicated that the arginine / glutamate mixtures were not appropriate in the buffer for NMR resonance assignment, as we initially believed. Given that buffers containing less than 50 mM salt resulted in biologically inactive enzyme, and that 20 mM Tris buffer containing 100 mM NaCl indicated no precipitation, CPG2₂₃₋₄₁₅ samples were prepared in this buffer for further NMR resonance assignment.

Table 3.3: CPG2 precipitation matrix to determine protein solubility and stability conditions for NMR studies using the X-ray screening method.

	(pH)	Sodium Citrate (5.5)	Na/K Phosphate (5.2)	Na/K Phosphate (6)	Na/K Phosphate (6.5)	Na/K Phosphate (7.1)	Tris (7.48)	Tris (8.05)	Tris (8.53)	HEPES (7.42)	MOPS (7)	PIPES (7.1)
50 mM	NaCl		*	*	*	*	*	*	*	*		*
	NaBr			*	*	*	*	*				*
	LiCl				*	*	*	*	*	*		*
	KCl			*	*	*	*	*	*	*	*	*
	KSCN			*	*	*	*	*	*	*	*	*
100 mM	Na ₂ SO ₄			*	*	*	*	*	*	*	*	*
	Arg+Glu	*		*	*	*	*	*	*	*	*	*
	NaCl			*	*	*	*	*	*	*		*
	NaBr			*	*	*	*	*	*	*		*
	LiCl			*	*	*	*	*	*	*		*
150 mM	KCl			*	*	*	*	*	*	*	*	*
	KSCN			*	*	*	*	*	*	*	*	*
	Na ₂ SO ₄			*	*	*	*	*	*	*	*	*
	Arg+Glu	*		*	*	*	*	*	*	*	*	*
	NaCl			*	*	*	*	*	*	*	*	*
200 mM	NaBr			*	*	*	*	*	*	*	*	*
	LiCl		*	*	*	*	*	*	*	*	*	*
	KCl			*	*	*	*	*	*	*	*	*
	KSCN			*	*	*	*	*	*	*	*	*
	Na ₂ SO ₄			*	*	*	*	*	*	*	*	*
	Arg+Glu			*	*	*	*	*	*	*	*	*

The asterisk indicates no precipitation of CPG2 in the corresponding condition after more than eleven days.

3.6 Preparation of Co-Crystals for X-ray Crystallography.

Due to its therapeutic applications in the treatment of cancer and autoimmune diseases, the three-dimensional structure of CPG2 is of great interest. Early X-ray crystallography investigation of CPG2 demonstrated that it consists of two domains, the catalytic domain which delivers the substrates for the two zinc ions in the active site, and the dimerisation domain (Figure 1.3) (Rowsell et al. 1997). This crystallographic data helped to better inform early rational drug design efforts (Khan et al. 1999), but further attempts at co-crystallisation of CPG2 with ligands, substrates or inhibitors have not been reported and very little progress has been made since these early studies to establish key features of CPG2. Hence we decided to attempt co-crystallisation of CPG2 with the most promising inhibitors CP06 and CP67 (see structures in Figures 4.1 and 4.2 of Chapter 4).

3.6.1 Protein Preparation.

X-ray crystallography requires formation of ordered crystals and much higher protein concentrations than those required in NMR studies, but one key difference is that protein samples for X-ray crystallography do not need isotopic labelling. So far, our sample preparation did not fail any of the criteria needed in protein preparation for crystallography. Pure samples of unlabelled CPG2₂₃₋₄₁₅ were obtained successfully (Figure 3.10-B), a high yield of soluble

protein was produced (Table 3.1), the protein was active (Figure 3.5-C) and highly stable. The only major difference was that the protein had to be brought into a far more concentrated state than in natural circumstances. As noted earlier, this fact is a rate-limiting step for NMR analyses as well. However, the rate of concentrating the sample differs for both techniques, and NMR requires less concentrating than X-ray crystallography.

Before beginning X-ray trials, the protein had to be prepared with buffer containing little to no salt. Since X-ray does not restrict from use of glycerol, commonly used as a cryo-protectant to increase protein stability, we decided to reduce the salt concentration to 50 mM in the optimised NMR buffer (20 mM Tris, 100 mM NaCl, 0.2 mM ZnCl₂, pH 7.3) and add 20% glycerol. In order to screen a range of conditions, 200 μ L of protein at 10-20 mg/mL, which was required and was easily achieved using centrifugal concentrators.

3.6.2 Co-Crystallisation Attempts.

Rowse et al. (1997) were the first to crystallise CPG2 (Rowse et al. 1997; Tucker et al. 1996). Hence, we decided to be guided by their final crystallisation protocol which involved mixing 4 μ L of crystallisation buffer solution [200 mM Tris-HCl pH 7.2; 12% (w/v) polyethylene glycol 4000, 200 mM zinc acetate, 10% (v/v) glycerol] with 4 μ L of 20.0 mg/ml protein solution to form a hanging drop which was allowed to equilibrate with the reservoir solution at room temperature (Tucker et al. 1996). For co-crystallisation with ligand, 10 molar excess of ligand was added to the protein sample after concentration.

A 24-well tissue culture trays was used to set up crystallisation screens using the hanging drop by vapour diffusion method (see Section 2.2.11 of Chapter 2). Firstly, the well was prepared and usually 1 ml of a buffered precipitant solution with different polyethylene glycol concentrations, pHs and glycerol content conditions was added (Table 3.4). Then 4 μ L of the concentrated protein sample was pipetted onto a coverslip, followed by 4 μ L of the well solution. The coverslip was then inverted over the well and sealed using grease. This was then left undisturbed for at least 24 hours to equilibrate at 18°C.

Table 3.4: First attempt at co-crystallisation: screening using the hanging drop method. Conditions were guided by Tucker et al. (1996). **X** indicates absence of crystals with all protein preparations, while **✓** indicates formation of crystals. Black checkmark (✓): crystals with CPG₂₃₋₄₁₅, red checkmark (✓): crystals with CPG₂₃₋₄₁₅ and glycosylated inhibitor (CP67), and green checkmark (✓): crystals with CPG₂₃₋₄₁₅ and inhibitor (CP06).

	10% Glycerol, 200 mM Tris			5% Glycerol, 200 mM Tris		
	pH 7.0	pH 7.2	pH 7.4	pH 7.0	pH 7.2	pH 7.4
8 % PEG 4000	X	X	X	X	X	X
10 % PEG 4000	X	X	X	X	X	X
12 % PEG 4000	X	X	X	X	X	X
14 % PEG 4000	✓✓✓	✓✓	✓	✓	✓✓	✓✓✓

At the start of the experiment, the precipitant concentration in the drop is half of that in the well. Equilibration then takes place via the vapour phase. Given the relatively large volume of the well, its concentration effectively remains the same. The drop however loses water vapour to the well until the precipitant concentration equals that of the well. If conditions are favourable the protein should become supersaturated and is driven out of solution in the

form of crystals. Crystallisation was unsuccessful below 14% PEG 4000. All the crystals produced were tested for diffraction (Figure 3.13). Model building² (where the electron density map is interpreted in terms of a set of atomic coordinates) was done using molecular replacement, as we already had a coordinate set to work with from Rowsell et al. (1997). Unfortunately, the ligands did not co-crystallise with the protein 3.4. This suggests, that the crystallisation buffer conditions used for co-crystallisation purposes were not ideal.

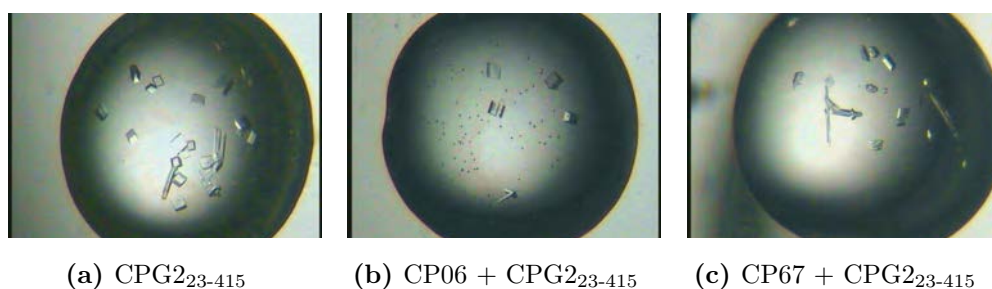


Figure 3.13: Crystal grown using conditions described in table 3.4. Crystals as observed under a microscope.

A second approach was then used in which a higher protein sample concentration (20-30 mg/mL) was mixed with with a larger ligand excess (30 and 40×molar excess), utilising commercially available plates JSCG+, PACT and Morpheus 96-well plates from Molecular Dynamics (Suffolk, UK) to screen for better buffer conditions for co-crystallisation (see Section 2.2.11 of Chapter 2). Protein began to crystallise as thin needles with 20 w/v PEG 6000, >100 mM NaCl and 0.1 M Tris (pH 7) buffer conditions (Figure 3.14).

²Dr Karen Ruane (School of Life Sciences, University of Warwick) tested crystals for diffraction, collected the data at Diamond Synchrotron Source in Oxford and solved the structure.



Figure 3.14: Thin needle of protein crystals. This needle grew in 0.2 M NaCl, 0.1 M Tris and 20% w/v PEG 6000 (pH7) buffer conditions.

Hence further buffer optimisation using 20% w/v PEG 6000 was thought to potentially generate appropriate co-crystals. A second crystallisation screen was set-up using the hanging drop by vapour diffusion method. The well was prepared with 1 ml of a buffered precipitant solution with 20% w/v PEG 6000 concentrations and various ligand-protein sample mixtures were added (Table 3.5). 1 μ L of the concentrated protein-ligand sample was pipetted onto a coverslip, followed by 1 μ L of the well solution.

Table 3.5 indicates that CPG2₂₃₋₄₁₅ crystallised in the presence of the glycosylated inhibitor (CP67), but failed in the presence of CP06 inhibitor. This could potentially be indicative of a weak binding between CP06 inhibitor and CPG2₂₃₋₄₁₅.

Table 3.5: Second trial at co-crystallisation: screening using the hanging drop method. ✗ indicates absence of crystals, while ✓ indicates formation of crystals. The ligand to protein ratio is indicated in the first column.

	20% PEG 6000					
	15%	17.5%	20%	22.5%	25%	27.5%
CP06 ligand, [40]:[1]	✗	✗	✗	✗	✗	✗
CP67 ligand, [30]:[1]	✗	✗	✗	✗	✓	✓

3.7 Conclusion.

At the start of the project, bacterially-expressed full-length CPG2 enzyme was highly insoluble. The intention was to redirect the enzyme from inclusion bodies to cytoplasmic fractions using the *E. coli* system, and produce soluble and active CPG2 enzyme for structural studies. This initially involved fine-tuning some of the expression factors such as changes in the *E. coli* expression strain and expression temperature, amongst others.

Different *E. coli* strains were tested for their ability to express CPG2, with a good yield, in soluble form in the growth media. The C41 (DE3) and BL21 (DE3) strain showed comparable expression of CPG2 at 20°C and similar induction concentration (Figure 3.1). In contrast, the C43 (DE3) strain did not survive in rich media (cell density OD₆₀₀ did not exceed 0.1). The yield, when using BL21 (DE3) as a host, was significantly improved upon use of a low induction temperature (15°C) in both rich and minimal media (Figure 3.2). Low IPTG concentrations were also tested, but had little impact on protein yield. Enzymatic activity of repeat batches of CPG2 preparation displayed high variability (Figure 3.3-A). SDS-PAGE gels revealed the presence of two bands of different density and thickness that migrate very close together (Figure 3.3-C). CPG2 gene has an N-terminal leader sequence that points the enzyme to the periplasm of *Pseudomonas* sp. Strain RS-16. Mass spectroscopy analysis of these two bands confirmed both bands were a match to full-length CPG2 protein (Figure 3.3-C, right panel). Not only did this indicate the production of two different isoforms of the same protein, which could be responsible for the batch-to-batch inconsistency observed

in activity assays, but the results also suggest that reducing the temperature of expression did not lead to removal of the signal peptide and subsequent soluble periplasmic CPG2 in *E. coli*. Thus, the N-terminal signal peptide was removed in a new construct (CPG2₂₃₋₄₁₅). This deletion resulted in the expression of one single protein with a 6-fold increase in protein yield (Figure 3.5-A and Table 3.1). CD measurements of full-length CPG2 and CPG2₂₃₋₄₁₅ indicated that CPG2₂₃₋₄₁₅ secondary structure values were in better conformity with those estimated from the crystal structure (Figure 3.5-D). The next phase was to overexpress ²H/¹³C/¹⁵N enriched *Pseudomonas* protein in *E. coli*.

Estimation of protein concentration via the BCA assay indicated over 50% reduction in soluble protein yield of CPG2₂₃₋₄₁₅ upon switching the growth medium from 2×YT to M9 with isotopic labels. Because cells containing the CPG2₂₃₋₄₁₅ expression plasmid grew well in rich medium, and poorly in labelled-minimal medium, methods proposed by Sivashanmugam et al. (2009) and Marley et al. (2001), which report double-colony selection in increasing amount of deuterated water and cell-condensation methods, respectively were ideal approaches to test. Double colony selection revealed variability in colony-to-colony unevenness in CPG2₂₃₋₄₁₅ expression levels and demonstrated the need to screen colonies for the best protein expression levels (Figure 3.9-A). Together with the condensation method, we managed to increase expression of labelled protein and reduce the cost of labelled-protein sample by four-fold, as compared to conventional labelling (Figure 3.9-B).

Successfully purified soluble CPG2₂₃₋₄₁₅ had limited long-term stability (activity declined 24 hours post-purification) and a propensity to aggregate

over-time. Consequently, different buffer conditions were tested and revealed that CPG2₂₃₋₄₁₅ activity was vastly affected by changes in salt concentration (Figure 3.11). The protein required >50 mM NaCl, which is known to increase sample conductivity and decrease sensitivity in NMR. We tested the low conductance buffer conditions used by Dötsch and Withers (2005), which consisted in using 50 mM arginine and glutamine instead of NaCl (Dötsch and Withers 2005). We were able to collect NMR spectra using this buffer condition. Unfortunately, the 0.5 mM protein sample started to precipitate after 2 days in the magnet, making 3D-experiments intractable for backbone assignment necessitating (3 days each) for this large protein. Thus, a typical screening technique used by crystallographers was applied to determine optimal sample conditions for structural NMR studies (Table 3.3). The 363 buffer conditions screened revealed that at pH values above 6, CPG2₂₃₋₄₁₅ remained soluble in all buffers for at least 5 days at 18°C, except when arginine and glutamate were included in the buffer. However, CPG2₂₃₋₄₁₅ remained soluble for 2 days in the presence of arginine and glutamate, which is consistent with what we observed previously. After 5 days, precipitation took place at pH values below 6, except when the concentration of NaCl was at or above 200 mM. Given that buffers containing less than 50 mM salt resulted in biologically inactive enzyme, and that 20 mM Tris buffer containing 100 mM NaCl indicated no precipitation, CPG2₂₃₋₄₁₅ samples were prepared in this buffer for further ligand-protein NMR studies (Chapter 4).

Moreover, expression optimisation of protein preparation and the quest for stable protein sample made crystal screening for X-ray crystallography all the

more applicable to this project. Although X-ray crystallography did not call for labelled protein as required for NMR analyses, it did require higher amounts of protein. Lowering the salt concentration and adding 20% glycerol made crystallisation achievable (Figure 3.13).

Chapter 4

Characterisation of Ligand Binding Properties using NMR Spectroscopy

4.1 Introduction.

Protein-ligand interactions are implicated in all biological processes. Generally the affinity can be characterised by the binding constant and / or the standard free energy. In this project, we were interested in the processes underlying the mechanism of recognition of CPG2 for its inhibitors, which were provided by our industrial partner Mologic Ltd. (Bedford, UK). Amongst a plethora of biophysical techniques, such as isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), X-ray crystallography, and mass spectrometry (MS), used to investigate protein-ligand interactions, NMR spectroscopy displays a

suitable soluble environment for the study of inhibitors targeting soluble proteins, such as CPG2. Generally, two approaches are utilised to study protein-inhibitor interactions when NMR spectroscopy is used. These two approaches are ligand-based and protein-based NMR experiments, in which NMR parameters of the ligand and the protein, respectively, are monitored and compared in their free and bound states (Cala et al. 2014). This chapter describes how NMR spectroscopy, using these two approaches, enabled us to provide information on the protein-inhibitor binding strength and specificity to CPG2.

4.2 CPG2 Inhibitors Used for NMR Studies.

Mologic Ltd.¹ (Bedford, UK) is using CPG2 as part of a fusion-protein with anti-cancer antibodies, so that it can convert a glutamate-conjugated pro-drug to a cytotoxic drug selectively on the surface of the tumour.

However, recent clinical trials justified the need for unbound CPG2 to be cleared from the circulation before administration of the pro-drug to prevent systemic toxicity in the patient. Hence, Mologic has developed peptide-based inhibitors that targeted the CPG2 enzyme. These peptides must also carry sugar residues that enable them to be bound by the asialoglycoprotein receptor of hepatocytes, to ensure efficient clearance of the inhibited CPG2 complex via the liver. As shown in Figures 4.1 and 4.2, these inhibitors are made of a

¹Dr James Schouten and Dr Joannah Towler (Mologic Ltd., Bedford, UK) produced the CPG2 inhibitors used in this study.

"warhead", which consists of an aromatic group (shown in black) and a glutamate moiety (shown in green), attached to a galactosylated tail (shown in red). The warhead is directly implicated in CPG2 recognition and its inhibition, whereas the galactosylated tail is present to ensure rapid removal of the inhibited CPG2 complex from the bloodstream.

Here inhibitors named CP06 and CP39 serve as two potential "warheads", while CP39 and CP42 are the two galactosylated versions, respectively. CP06 and CP39 differ only by the type of aromatic ring (represented in black in Figures 4.1 and 4.2 respectively) present in their structure: CP06 has a naphthalene ring connected to a glutamate moiety, whereas CP39 has an homophenylalanine that has an additional CH₂ group compared to phenylalanine (for simplification we will refer to it as the benzyl group). Inhibitors CP67 and CP42 further differ in structure by their galactosylated tail. CP67 has three serine galactose groups, when CP42 only contains one. To better understand the mode of binding of these peptide-based ligands, Mologic is interested in acquiring structural and thermodynamic data for the free enzyme and ligand-enzyme complex using solution-state nuclear magnetic resonance spectroscopy (NMR). Moreover, we are interested in the differences in the affinity and molecular recognition between the inhibitors and their glycosylated structures. CP06 and CP67 retained inhibition activity, while one of the inhibitors (CP39) lost its aptitude to inhibit CPG2 upon galactosylation (i.e CP42). For the purpose of this project, we wanted to analyse these two pairs of structures to understand the structural features underlying selective CPG2 inhibition. Figures 4.1 and 4.2 shows all the different inhibitors we

wanted to test. Unfortunately, only the structures displayed in bold were analysed (i.e. CP06, CP67 and CP42), as CP39 degraded according to the 1D ^1H NMR data we recorded.

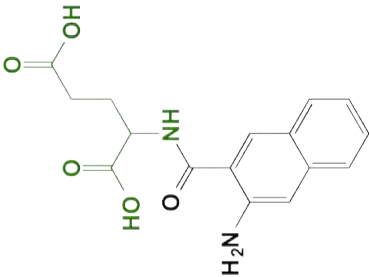
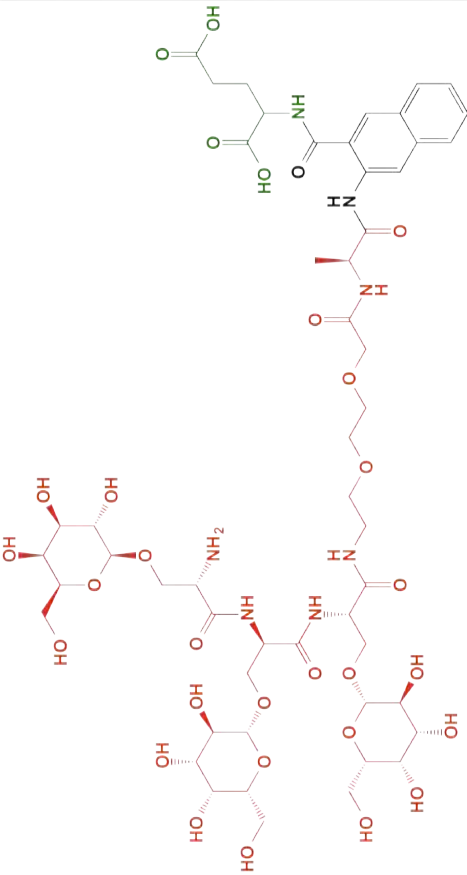
Name	Compound Structure	Structure	Molecular Weight, Da	Inhibitory constant against CPG2 (Octet RED96 System)
CP06		Amino-Naphthoic Acid (ANA)-Glutamate (Glu)	316.31	IC ₅₀ = 88.5 μM
CP67		Serine-Glactose (SerGal) α-ANA-Glu	1280.20	IC ₅₀ = 93 μM

Figure 4.1: CP06 and CP67 Inhibitors produced by Mologic Ltd. These inhibitors were provided by Mologic (Bedford, UK) to study their mechanism of interactions with CPG2. The IC₅₀ was obtained using the Octet RED system with immobilised CPG2 onto the sensor chip.

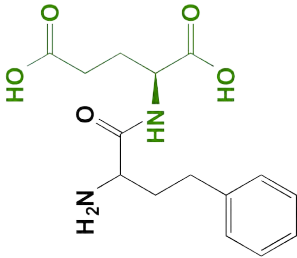
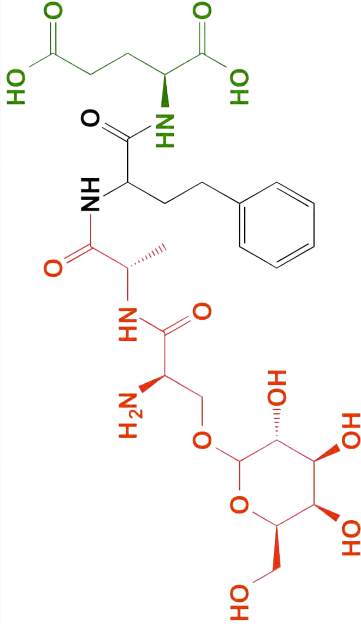
Name	Compound Structure	Structure	Molecular Weight, Da	Inhibitory constant against CPG2 (Octet RED96 System)
CP39		Homophenylalanine (HPA)-Glu	308.33	IC ₅₀ = 61 μM
CP42		SerGal-α-Ala-HPA-Glu	628.63	IC ₅₀ = No significant inhibition

Figure 4.2: CP39 and CP42 Inhibitors produced by Mologic. These inhibitors were provided by Mologic (Bedford, UK) to study their mechanism of interactions with CPG2. The IC₅₀ was obtained using the Octet RED system with immobilised CPG2 onto the sensor chip.

4.3 Ligand-Based NMR Experiment: STD-NMR.

1D-NMR experiments were used to probe the mechanisms of recognition of these three inhibitors to CPG2 in solution to determine whether the binding regions of the inhibitor were conserved between the inhibitors and their matching galactosylated inhibitors. We wanted to know if any particular functional group was specifically accountable for inhibition. CP06, CP67 and CP42 inhibitors were hence tested with CPG2 using saturation transfer difference (*STD*) NMR experiment.

To obtain information about the binding specificity at atomistic level, STD-NMR was used (Cutting et al. 2007). For the purpose of describing the process, CP06 inhibitor has been used as an example (results for all three inhibitors are shown in Section 4.3.3). Figure 4.3 reveals the different steps involved in the STD-NMR experiment used for CP06. Spectrum A shows a ^1H -NMR spectrum of the free CP06 ligand in solution. STD-NMR relies on the transfer of saturation from the protein to ligand. Hence, spectrum B in Figure 4.3 corresponds to the spectrum of the same sample recorded with off-resonance irradiation applied at 38 ppm, far from the on-resonance frequency, and protein and ligand NMR resonances, so that no NMR signals are modified (Viegas et al. 2011). Spectrum C was recorded with on-resonance irradiation at -0.5 ppm, where only protein signals are irradiated. Finally, spectrum D shows the difference spectrum resulting from the subtraction of the off-resonance and on-resonance spectra. This spectrum only shows resonances that have

undergone saturation, namely the protein and the bound ligand resonances. Because the protein is present at low concentration and has a high molecular weight, the protein signals will relax too fast to be visible and the result is a simple 1D ^1H spectrum that reveals only the binding nuclei. Moreover, impurities observed in spectra B are no longer detectable in spectrum D, hence the impurities contained in the spectrum are effectively subtracted and therefore do not give rise to signals in the difference spectrum.

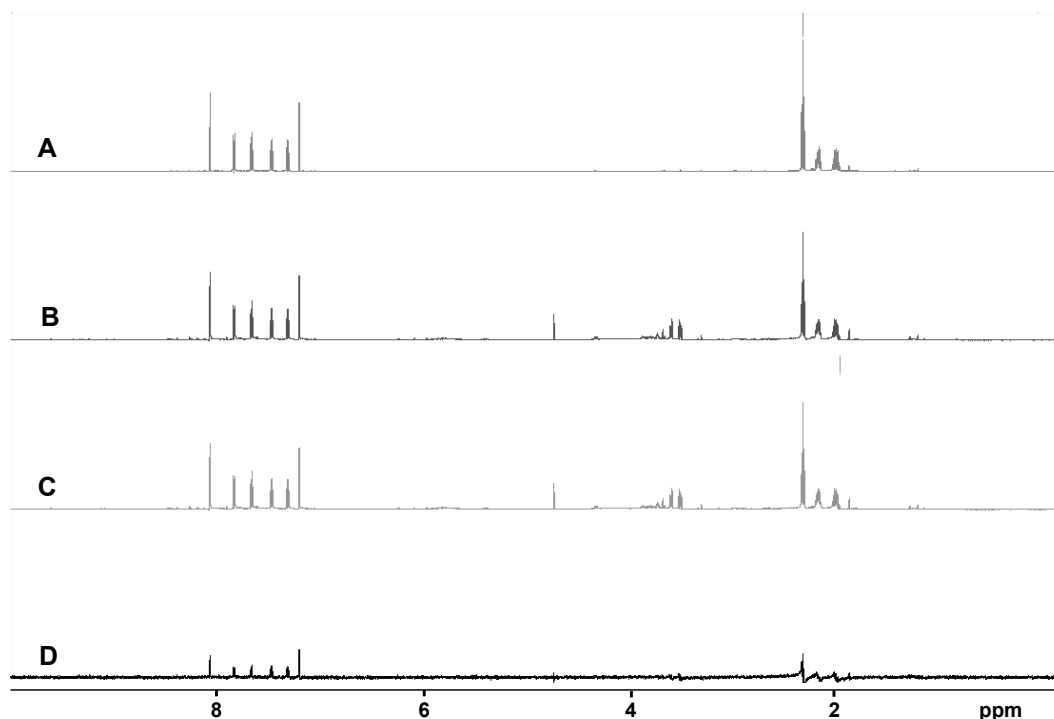


Figure 4.3: STD-NMR: 3 step concept. *A)* Reference 1D ^1H NMR spectrum of the CPG2 and CP06 sample mixture. *B)* Corresponding off-resonance spectrum with irradiation at 38 ppm. *C)* Corresponding off-resonance spectrum with irradiation at -0.5 ppm. *D)* Corresponding STD-NMR spectrum showing that CP06 yields signals and therefore binds to CPG2.

For this work, the protein must be selectively saturated. This is possible because protein resonances have significant intensity even in the upfield ppm region from 0 to -1 ppm, where the preponderance of methyl NMR resonances of folded proteins occur and no ligand signal occurs or above the downfield

region above 10 ppm, hence outside the spectral window of the low-molecular weight inhibitors (Wuthrich 1998). Transfer of magnetisation within the protein, via spin diffusion, leads to fast saturation of the entire protein and not only of the few directly saturated resonances (Kalk and Berendsen 1976). Furthermore, for the on-resonance experiment to be successful, any direct irradiation of ligand frequencies needs to be avoided (false-positives). This can be rapidly checked using a ligand-only sample (in the absence of protein), and running an STD-NMR experiment. Figure 4.4 reveals that in the absence of the protein, the ligand is directly saturated at 0 ppm (Figure 4.4-A) and -1 ppm (Figure 4.4-C), but not at -0.5 ppm (Figure 4.4-B). This reveals the importance of screening beforehand for the appropriate on-resonance frequency in order to prevent false-positives. The on-resonance value for the other two ligands was set at -1 ppm using the same procedure, as -0.5 ppm revealed direct saturation of the ligand signals.

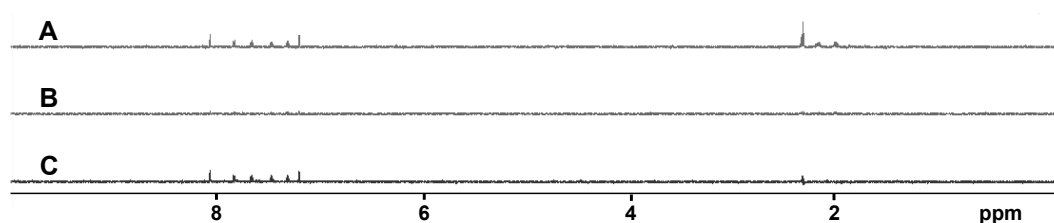


Figure 4.4: Testing on-resonance frequencies. To prevent the presence of false-positive signals due to direct irradiation of ligand protons, different on-resonance frequencies at *A*) 0 ppm, *B*) -0.5 ppm and *C*) -1 ppm were investigated using a ligand-only sample. Results showed that ligand signals were directly saturated at 0 and -1 ppm, only -0.5 ppm displayed the absence of false-positive signals.

4.3.1 Use of D₂O in Sample Preparation is Essential.

Before measuring the samples in 100% D₂O, samples in 90% H₂O and non-deuterated Tris buffer were used. Unfortunately, although the water suppression pulse program efficiently suppressed water signal around 4.701 ppm, negative NOE signals were observed in the downfield aromatic region (Figure 4.5-A). Non-binding inhibitors, which should only have contact to "unbound" water molecules found in solution, can be detected with a positive NOE because inhibitors that interact with bulk water will experience much faster tumbling (Ludwig and Guenther 2009). In this case, the negative NOE was originating from "bound" water molecules, i.e within the binding site of the target protein, where bound ligand directly interacted with "bound" water molecules via water-ligand-protein or protein-ligand complexes, whose rotational correlation time yields to negative cross-relaxation rates, thus negative NOE with water (Kieffer et al. 2011). This pathway of saturation can be overlooked if the ligand signals were to be detected using the water ligand observation with gradient spectroscopy (*WaterLOGSY*) NMR experiment (Dalvit et al. 2001). This effect can also be minimised when the samples are measured in 100% D₂O and if protein saturation is performed far from the water signal at 4.701 ppm (Figure 4.5-B). The latter was used as ligands were easily soluble in D₂O and the ligands of interest had NMR signals in the 2 to 8 ppm region of the spectrum, with very little chance to take part in protein saturation around the water signal region for STD-NMR experiment (Dalvit et al. 2001).

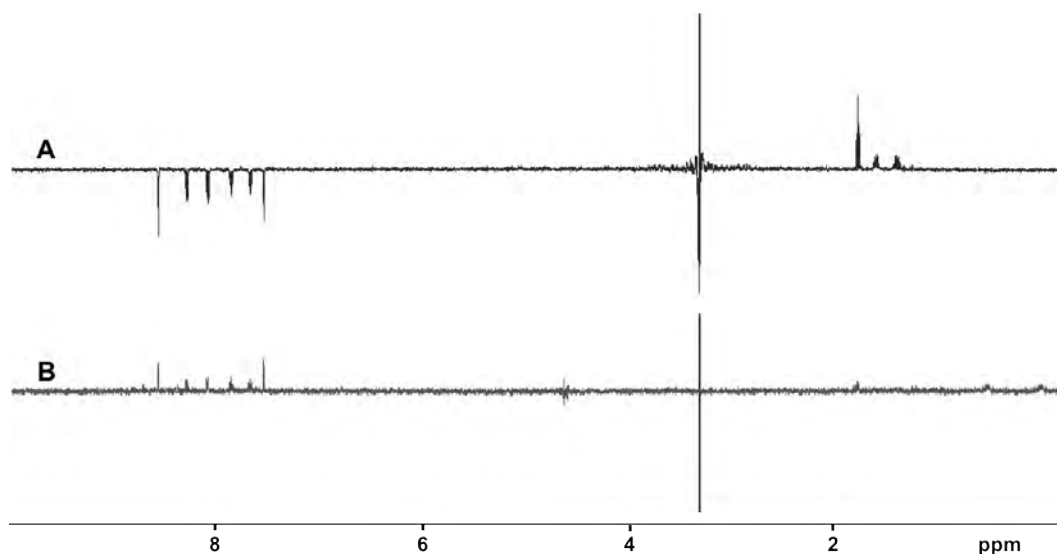


Figure 4.5: Buffer conditions for STD-NMR experiments. Two buffers were tested: *A*) a 10% deuterated buffer and *B*) a 99% deuterated buffer. 10% deuterated buffer showed presence of negative NOE downfield due to the presence of non-binding inhibitors in solution interacting with "bound" water molecules at the active site. Hence, for all STD-NMR experiment 99% deuterated buffer was used.

4.3.2 Saturation Dependence of the STD Intensities.

To investigate the influence of the saturation time on the STD effect, a set of six STD-NMR experiments was collected where concentrations of protein and ligand were fixed at 20 μM and 2 mM respectively. The saturation times were: 0.5 s, 0.75 s, 1 s, 1.25 s, 1.5 s and 2 s.

The time-course of saturation was obtained by plotting the STD amplification factor against the saturation time. Figure 4.6 shows the saturation profile for some selected resonances of the CP06 ligand at one ligand concentration (2 mM). The STD-AF factor of the aromatic protons (H6 and H4 of CP06 and CP67 respectively) were larger than that of the rest protons. Moreover, one can observe that longer saturation times (i.e 2 s) and high ligand concentration were required to reach the maximum observable STD effect. The

relative STD intensities, which were specified for CP06 in Figure 4.6-A, were therefore specific for a given saturation time and ligand concentration (2 mM) and cannot serve as a final value. Moreover, CP67 build up curves in Figure 4.6-B suggested that lower saturation time results in very similar STD-AF percentage amongst the various protons, prohibiting from experiencing differences between these different protons. High saturation times allowed a good differentiation between the STD-AF of the different protons. Hence, a 2 s irradiation time was used for epitope mapping of the inhibitors.

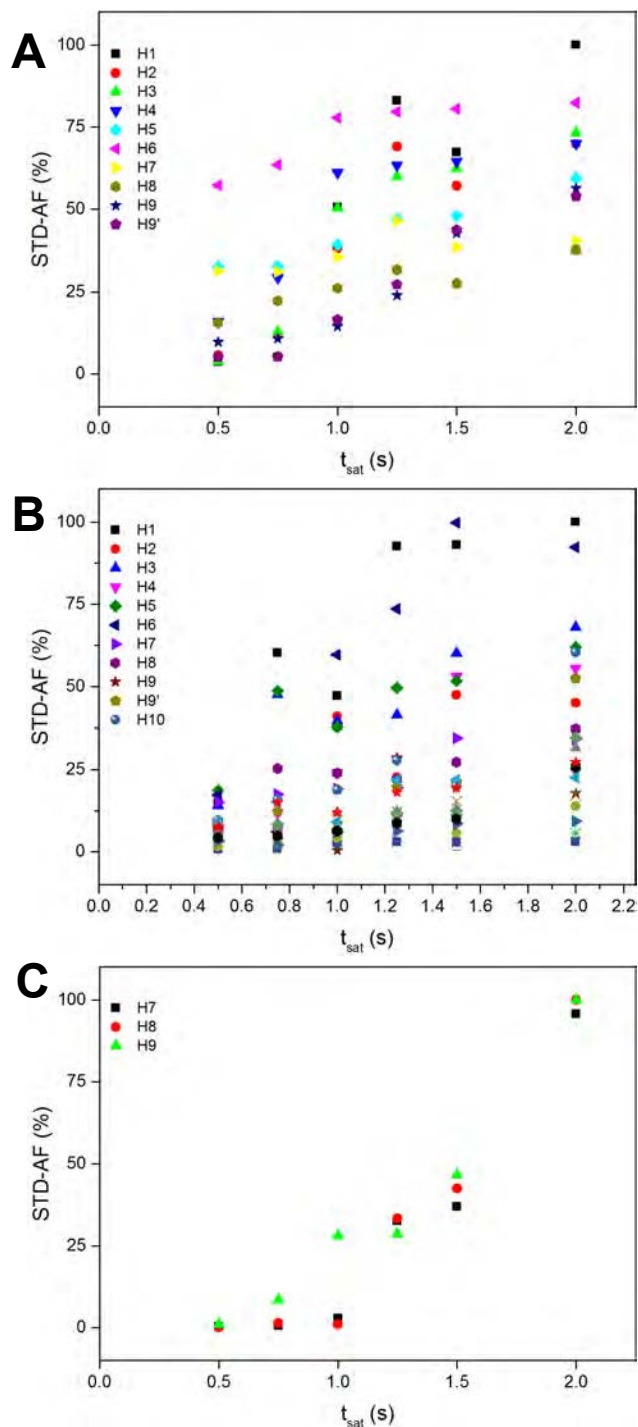


Figure 4.6: Saturating time t_{sat} build-up curves. Saturation transfer from CPG2 to A) CP06, B) CP67 and C) CP42 during STD-NMR experiment at 100:1 ligand to protein concentration ratio. Build-up curves obtained for different protons, protons of the ligand that are close to the protein, see Figures 4.7, 4.8 and 4.9 for assignment) are represented in different colours and given within each panel.

4.3.3 Ligand Group Epitope Mapping (GEM).

Figure 4.7 shows the reference at the top ^1H -NMR spectrum of CP06 with CPG2 in a 100:1 molar ratio and the corresponding STD NMR spectrum at the bottom. The STD spectrum indicated that CP06 binds to CPG2, since large signals of the ligand were observed. In addition, protons of the ligand involved in interactions with the protein were identified from the STD NMR spectrum because they were saturated to the highest degree. The difference signal intensities of the individual protons were normalised to the integral values of the highest signal of CP06, H3, set to 100%. Figure 4.7 shows the relative degree of saturation of each individual proton. The H1, H2, H3, H5 and H6 aromatic protons all have similar STD intensities between 89% and 100%. On the other hand H8, the only other proton displaying saturation, had the smallest intensity at 78%. Thus a clear distinction between protons with a strong contact to the protein, i.e. the aromatic protons, and the others, was made. Presence of saturation for H8 confirmed that glutamate is interacting with CPG2 and is presumably required for the CP06 warhead to be recognised by CPG2.

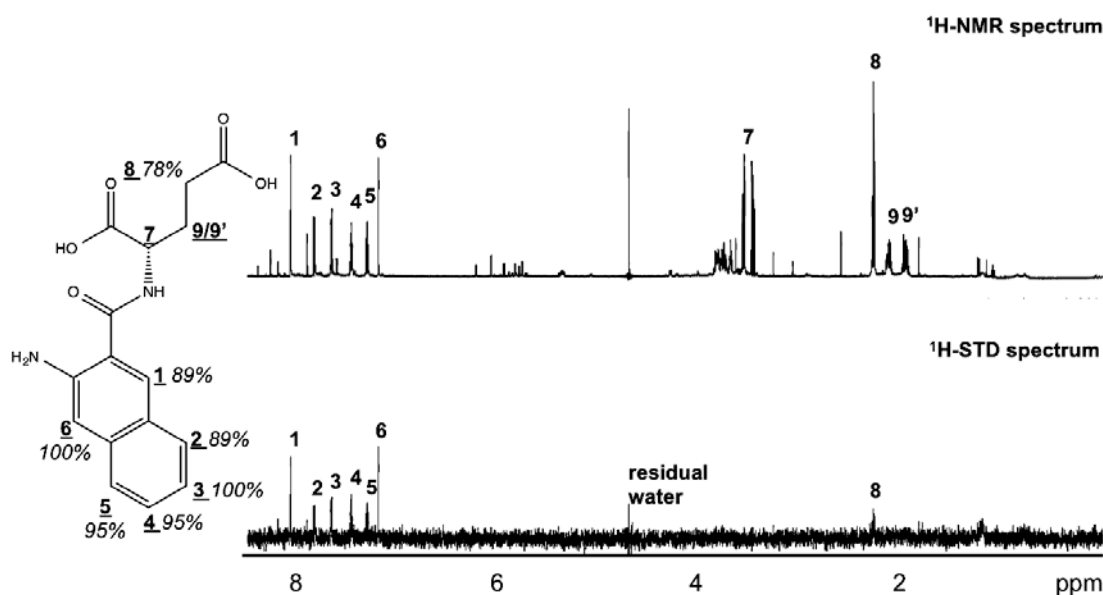


Figure 4.7: STD NMR spectrum of mM CP06 in the presence of 20 μ M CPG2 recorded at 298 K on a 700 MHz spectrometer. STD intensities relative to the corresponding reference intensities are shown next to each proton in the ligand (the assignment of each signal is indicated with an underlined number). The percentages shown on the structure are the STD epitope map after normalising the largest STD intensity to 100%. The small underlined numbers on the structure are the atom identification. The pre-saturation time of STD experiment was 2 s with a total recovery delay of 4 s. CPG2 was irradiated with a Gaussian shaped pulse at -0.5 ppm. A CMPG T2 filter of 2.5 ms was used to eliminate the protein background. The reference spectrum was recorded by irradiating at 38 ppm.

Investigation of the CP67 ligand (i.e. the galactosylated CP06 warhead) is displayed in Figure 4.8, with top spectrum being the reference NMR spectrum of CP67 with CPG2 in a 100:1 ligand to protein concentration ratio and bottom spectrum the corresponding STD-NMR spectrum. The STD spectrum illustrates CP67 binding to CPG2, since large signals of the ligand resonances were observed. In addition, protons of the ligand which are nearest to proton of the protein, were again easily identified from the STD NMR spectrum, because they were saturated to the highest extent. The different signal intensities of the individual protons were best analysed from the integral values of the highest signal of CP67, H3, set to 100%. Figure 4.8 demonstrated the relative degree of saturation of each individual protons. The H1, H2, H3, H5 and H6 aromatic protons all have similar STD intensities between 89% and 100%. On the other

hand H7, H8 and H9/9', had the smallest intensities, ranging from 58% to 78%, whereas the galactosylated tail showed no apparent proximity to the target protein, leaving it accessible to the hepatocyte receptors for clearance. Thus an evident distinction between protons with a strong contact to the protein (i.e. the aromatic protons) and the others, was mapped out. Since none of the protons from the galactosylated region of the inhibitor seemed to be in close to CPG2, the protons were not assigned on the ^1H -NMR spectrum. The results supported the hypothesis that the naphthalene ring plays a significant role in the molecular recognition of the inhibitor by CPG2. Likewise, it corroborates the requirement of a hydrophobic pocket near the CPG2 active site. With CP67 the glutamate group was again in close proximity to CPG2, as the glutamate protons were irradiated. Therefore, these results reinforced the idea that glutamate is evidently required for the inhibitor to be recognised by CPG2.

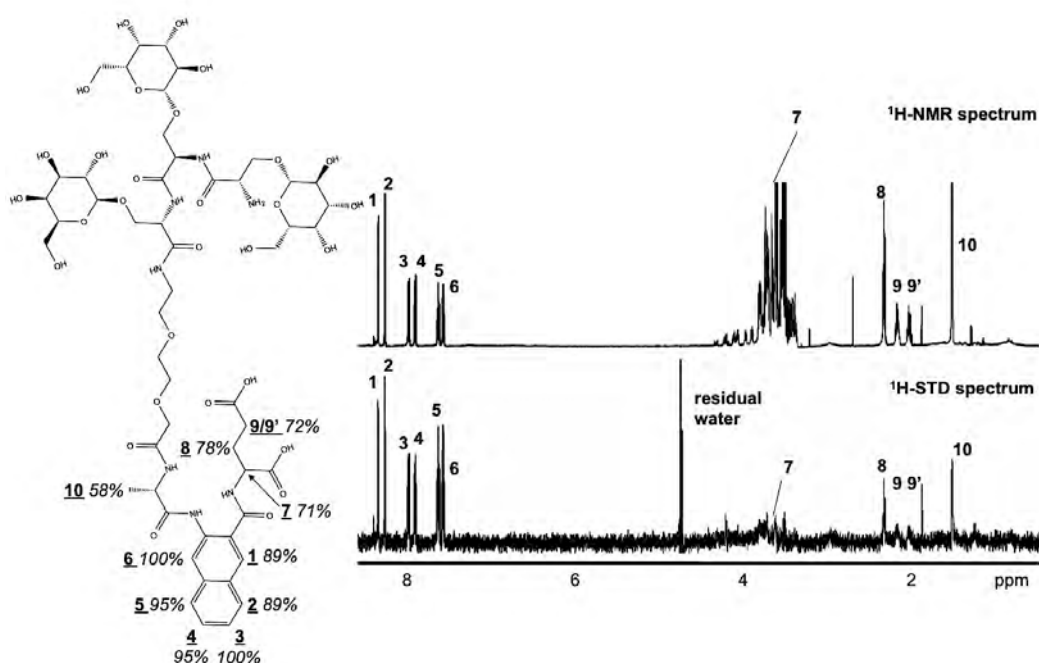


Figure 4.8: STD NMR spectrum of mM CP67 in the presence of 20 μM CPG2 recorded at 298 K on a 700 MHz spectrometer. STD intensities relative to the corresponding reference intensities are shown next to each proton in the ligand (the assignment of each signal is indicated with an underlined number). The percentages shown on the structure are the STD epitope map after normalising the largest STD intensity to 100%. The small underlined numbers on the structure are the atom identification. The pre-saturation time of STD experiment was 2 s with a total recovery delay of 4 s. CPG2 was irradiated with a Gaussian shaped pulse at -1 ppm. A CPMG T2 filter of 2.5 ms was used to eliminate the protein background. The reference spectrum was recorded by irradiating at 38 ppm.

Results for CP42 are given in Figure 4.9, with top spectrum being the reference ^1H -NMR spectrum of CP42 with CPG2 in a 100:1 molar ratio, and bottom spectrum the corresponding STD-NMR spectrum. The STD spectrum indicates that CP42 binds to CPG2, since large signals of the ligand can be observed. In addition, protons of the ligand which are closest to CPG2 can be once more easily identified from the STD-NMR spectrum, because they are saturated to the maximum degree. The difference signal intensities of the individual protons are best analysed from the integral values of the highest signal of CP42, H8, set to 100%. Figure 4.9 displays the relative degree of saturation of each individual proton. This time H10 displays the lowest STD intensity. On the other hand H7, H8, H9, display a saturation intensities ranging from 69% to 100%. Thus a clear distinction between protons with a strong contact to the protein, i.e. the protons found in the galactosylated region, and the others, found in the glutamate / benzyl region can be made. Although, the benzyl ring is in close proximity to the CPG2 binding site, the absence of saturated glutamate protons suggests that the glutamate is not interacting with CPG2. This shows that the type of aromatic ring used and the variation in galactosylation may play an important role in the molecular recognition of the inhibitor by CPG2, as CP42 contains a different aromatic ring and galactosylation than its predecessors, CP06 and CP67, which uses a naphthalene ring and three galactoses.

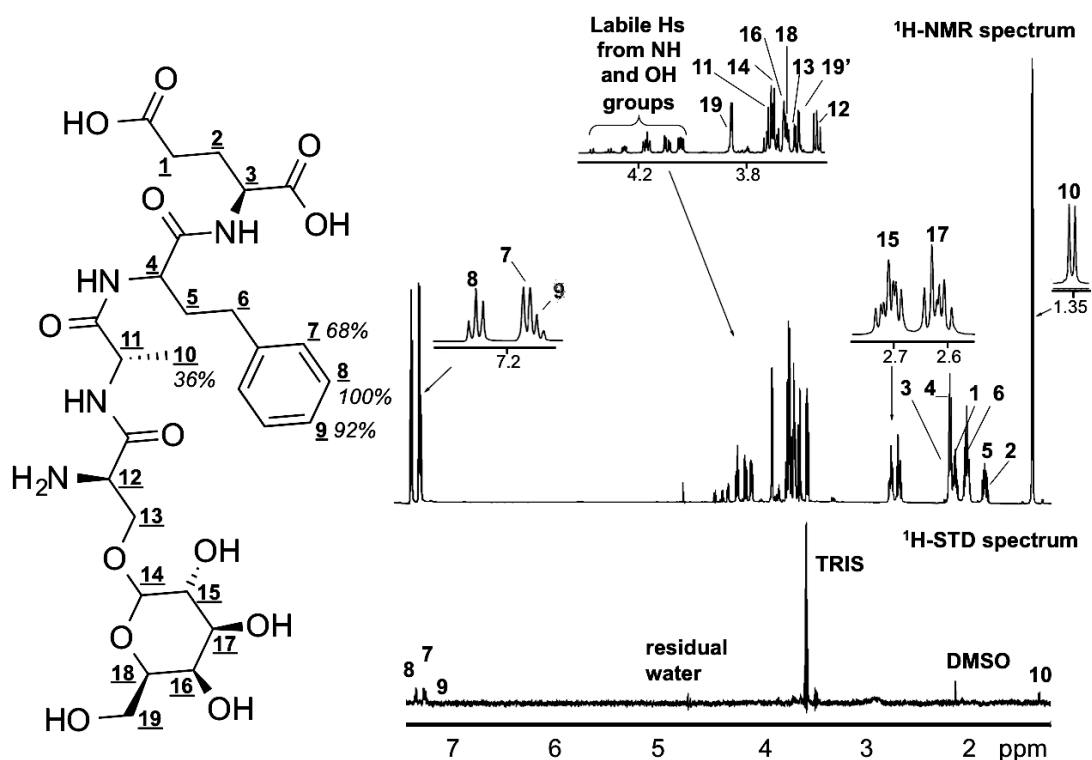


Figure 4.9: STD NMR spectrum of mM CP42 in the presence of 20 μM CPG2 recorded at 298 K on a 700 MHz spectrometer. STD intensities relative to the corresponding reference intensities are shown next to each proton in the ligand (the assignment of each signal is indicated with an underlined number). The percentages shown on the structure are the STD epitope map after normalising the largest STD intensity to 100%. The small underlined numbers on the structure are the atom identification (the assignment of the galactosylated group was not made, as it did not lead to any NOE saturation). The pre-saturation time of STD experiment was 2 s with a total recovery delay of 4 s. CPG2 was irradiated with a Gaussian shaped pulse at -1 ppm. A CPMG T2 filter of 2.5 ms was used to eliminate the protein background. The reference spectrum was recorded by irradiating at 38 ppm.

4.3.4 Better Assessment of the Absolute Magnitude of the Saturation Effect.

STD-NMR data can be related to K_D because spectra only contains signals from the bound ligands and, therefore, report information about the protein-ligand affinity. As with all the procedures used to study ligand-protein interactions, a titration with ligand is conducted to plot the response as a function of the initial ligand concentration.

The concentration of CPG2 protein used was 20 μM and ligands were titrated onto the corresponding protein samples from concentrated stock solutions (Figure 4.10). Each time y μL of ligand was added, y μL of 40 μM solution of CPG2 was added in order to keep the concentration of CPG2 constant throughout the titration. For titration and saturation time experiments, the STD amplification factor (*STD-AF*) was used (Heller and Kessler 2001; Angulo and Nieto 2011). This factor allowed a better visualisation of the complete extent of the saturation effect. The STD-AF is the fractional saturation of a given proton multiplied by the excess of the ligand over the protein. In effect, STD-AF is the relative intensity of the STD signal compared to that of a signal of the protein, i.e. STD-AF of 13 means that the STD signal has an intensity 13 times higher to that of a CPG2 proton. It provides an easy measure to quantify the amplification of the protein information observed in the STD signals of the ligand. Evidently, the STD-AF can be higher than 1 only if the residence time of a ligand is significantly shorter than the saturation time and if the concentration is much higher than that of the protein (Angulo and

Nieto 2011). Hence the STD-AF was always normalised to the signal displaying the highest STD-AF value.

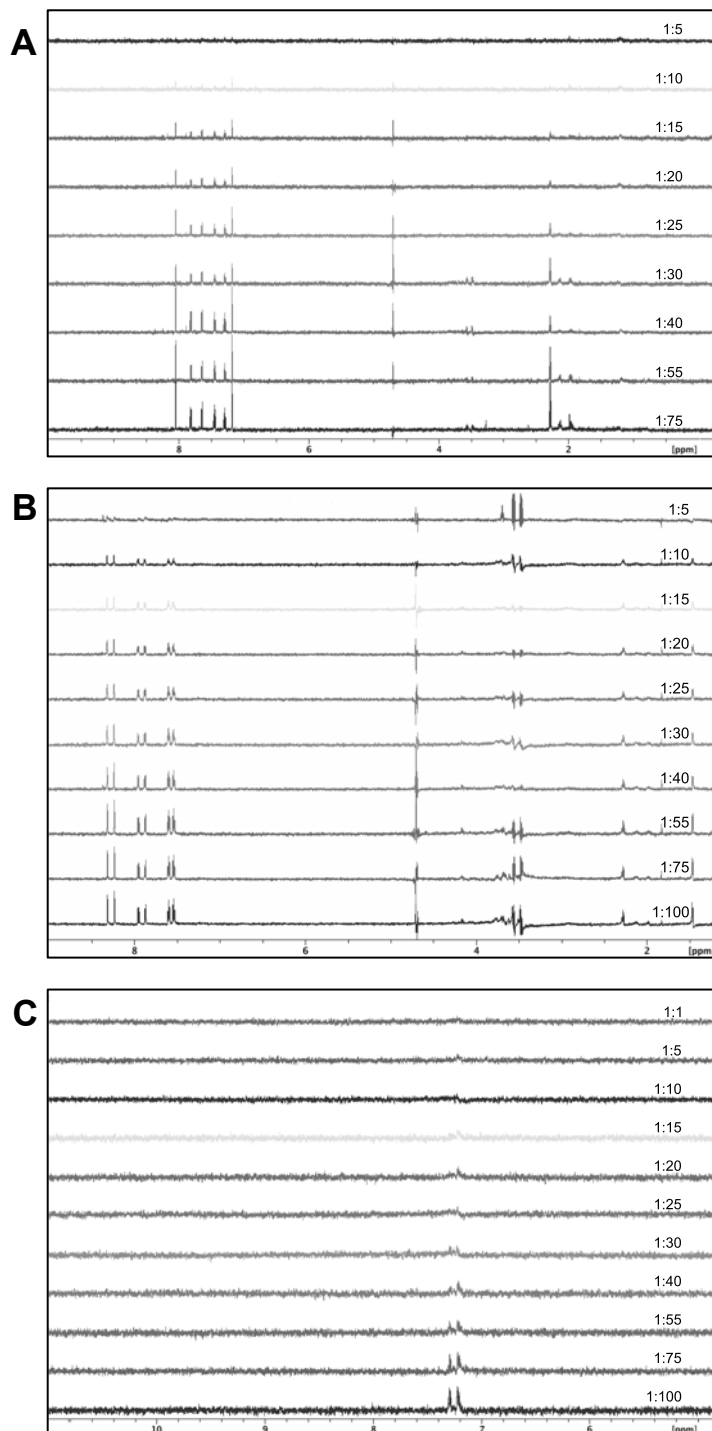


Figure 4.10: STD-NMR titration experiments of CPG2 by inhibitors. Titration using A) CP06, B) CP67 and C) CP42 inhibitors. The protein to ligand concentration ratios are displayed on the left, on top of each STD spectra. As titration progressed, the extent of saturation was increasing for the signals that were close to CPG2 surface.

The equation $(I_o - I_{\text{sat}})/I_o$ is the fractional STD effect expressing the signal intensity in the STD spectrum as a fraction of the intensity of an unsaturated reference spectrum. Undoubtedly, the fraction of ligands which are saturated is continuously reduced when the ligand excess is increased (Goldflam et al. 2012). By multiplication of the ligand excess with the fractional STD effect the STD-AF is obtained, as described by Equation (4.1).

$$STD - AF = \frac{(I_o - I_{\text{sat}})}{I_o} \times \frac{[L]}{[P]} \quad (4.1)$$

In this equation, I_o is the intensity of one signal in the off-resonance NMR spectrum, I_{sat} is the intensity of one signal in the on-resonance NMR spectrum, and $I_o - I_{\text{sat}}$ represents the intensity of one signal in the STD-NMR spectrum.

Figures 4.11, 4.12 and 4.13 show the binding isotherms obtained by plotting the STD-AF values of each proton that resulted in an NOE against the ligand concentration. For each proton, STD-AF build up curves were recorded (Angulo and Nieto 2011). The binding isotherms of CP06 protons, H7, H8 and H9/9' curves did not converge at high ligand concentrations, suggesting that the glutamate moiety weakly bound to CPG2, although they were in close proximity to it. For CP67, we see an identical pattern, where only protons 1 to 10 located in the "warhead" domain were directly involved in binding to CPG2. On the other hand for CP42, the second glycosylated inhibitor, only the aromatic protons showed NOEs. Since the curves did not converge at high concentration, it indicated these protons are only weakly binding to CPG2, which is in close agreement with the IC_{50} observed by Mologic.

Plotting these values at increasing values of ligand concentrations yields to a Michaelis-Menten hyperbolic dose-response curves described by Equation (4.2) (Angulo and Nieto 2011).

$$STD - AF_{[L]} = \frac{STD - AF_{\max} \times [L]}{([L] + K_D)} \quad (4.2)$$

where $STD - AF_{\max}$ is a dimensionless scaling factor representing the maximum STD amplification factor for the monitored signals. For each ligand concentration, an STD-spectrum was recorded. A plot of the STD-AF against ligand concentration was then used to measure the dissociation constant of the protein-ligand complex (Angulo and Nieto 2011).

STD amplification factor titration curves from some of the ligand protons are presented in Figures 4.11, 4.12 and 4.13). CP06 H1 to H6 protons binding isotherms fitted nicely to the Michaelis-Menten Equation (4.2), whereas H8 and H9 did not converge 4.12. This was indicative of very weak binding from H8 and H9, but suggested close interaction of these glutamate protons to CPG2.

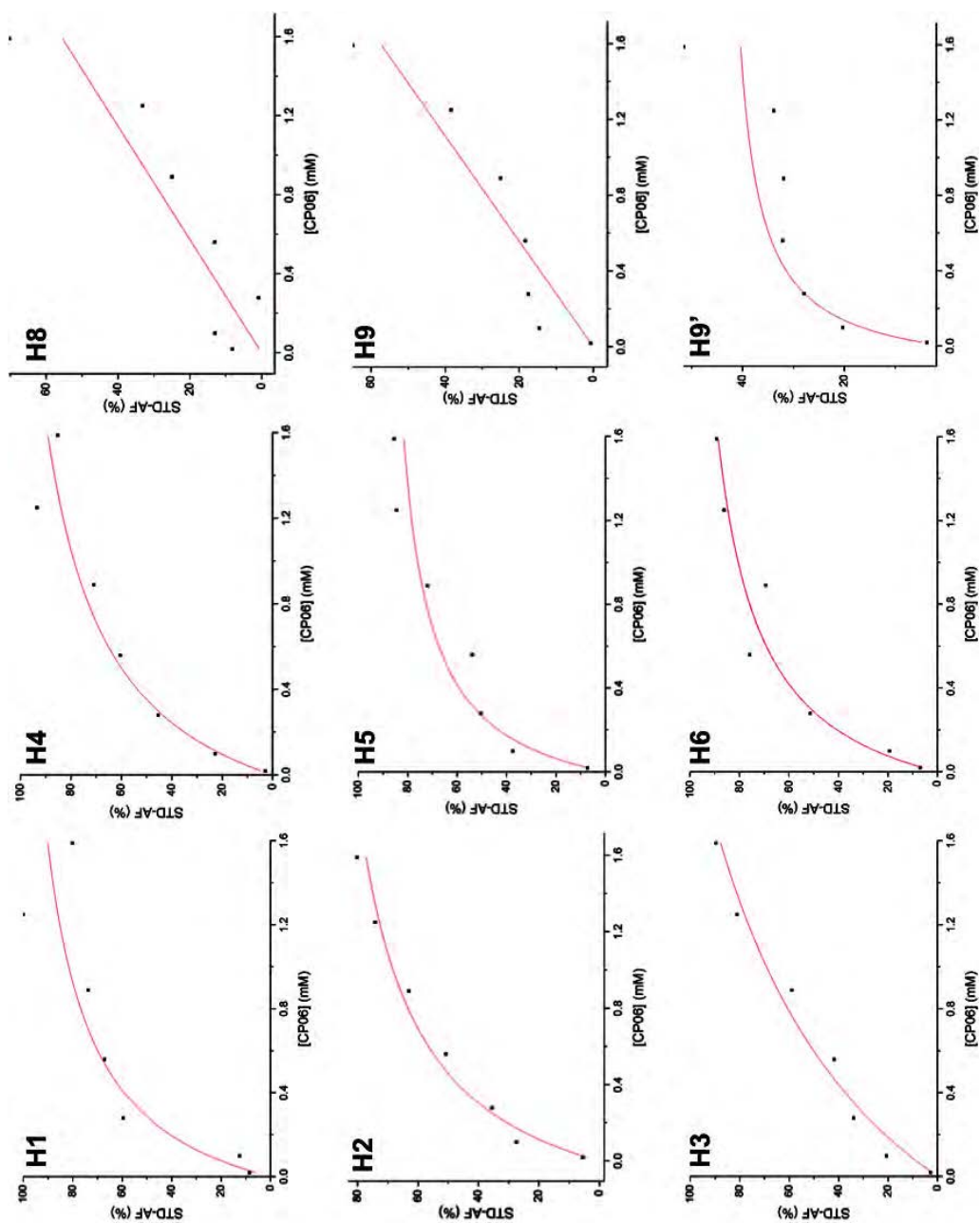


Figure 4.11: Binding isotherms fits for CP06. The STD-AF growth curve of each CP06 protons exhibiting saturation (dots), and the fit to Equation (4.2) (red line).

CP67 H1 to H6 protons binding isotherms also fitted to Michaelis-Menten equation, with very similar curves to that of CP06 aromatic protons (Figure 4.12). This suggested, again, that the naphthalene was required for CPG2 recognition, binding and possibly inhibition. Maximum STD-AF for H7 to H10 was reached very rapidly, indicating very strong binding from the glutamate group. This also implied that the glutamate moiety was indispensable for recognition and binding to CPG2. As opposed to CP06, protons that experienced saturation, all demonstrated high affinity to CPG2. These findings validated the inhibitory constants observed for CP67 and CP06 using the octet RED system (Figure 4.1).

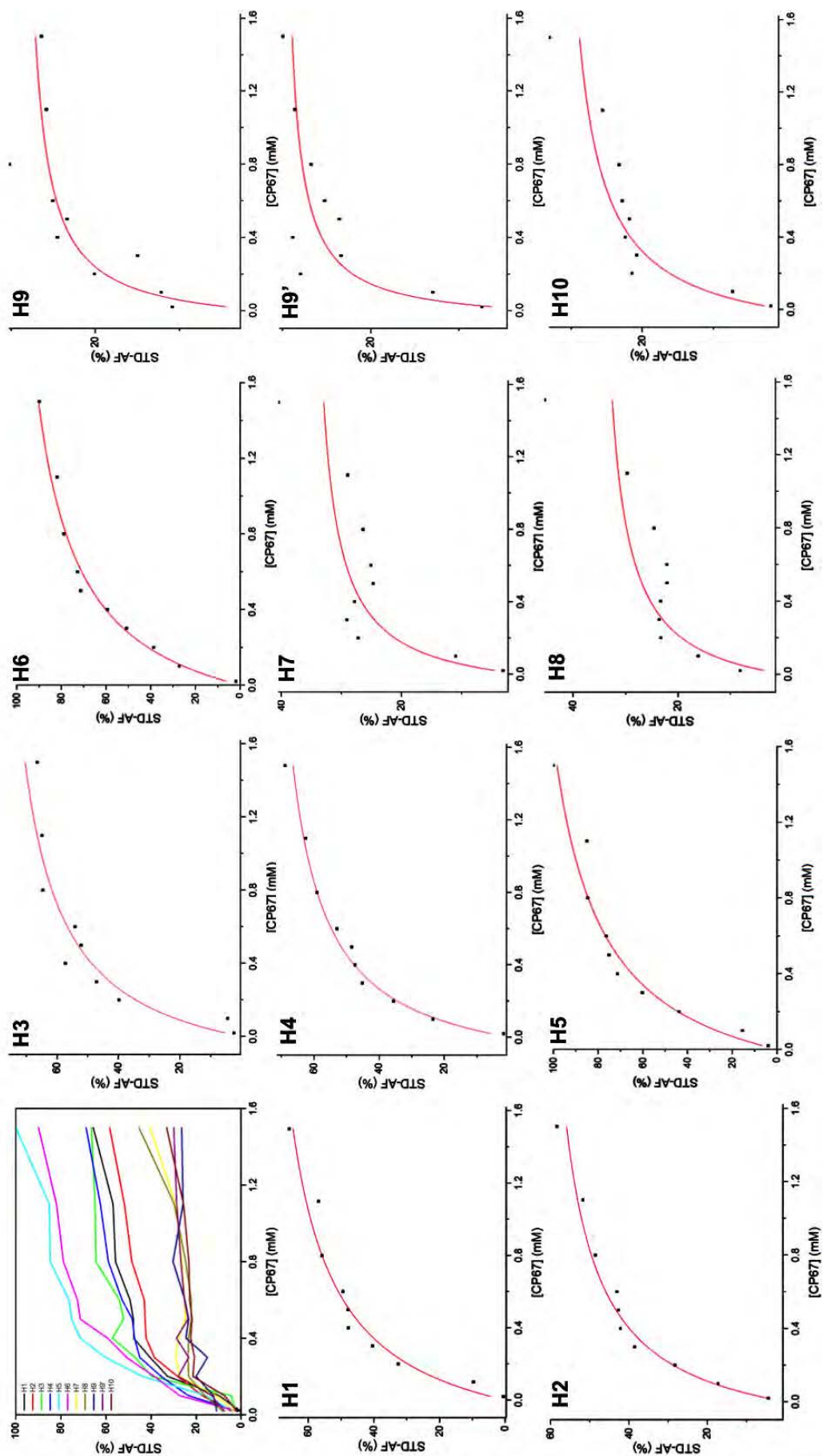


Figure 4.12: Binding isotherms fits for CP67. The STD-AF growth curve of each CP67 protons exhibiting saturation (dots), and the fit to Equation (4.2) (red line).

As shown in Figure 4.13, CP42 binding isotherms suggested very weak binding to CPG2 as STD-AF saturation was not reached, not even at 80:1 ligand to protein concentration ratio. This substantiated the absence of inhibition reported by Mologic using the octet RED system (Figure 4.2).

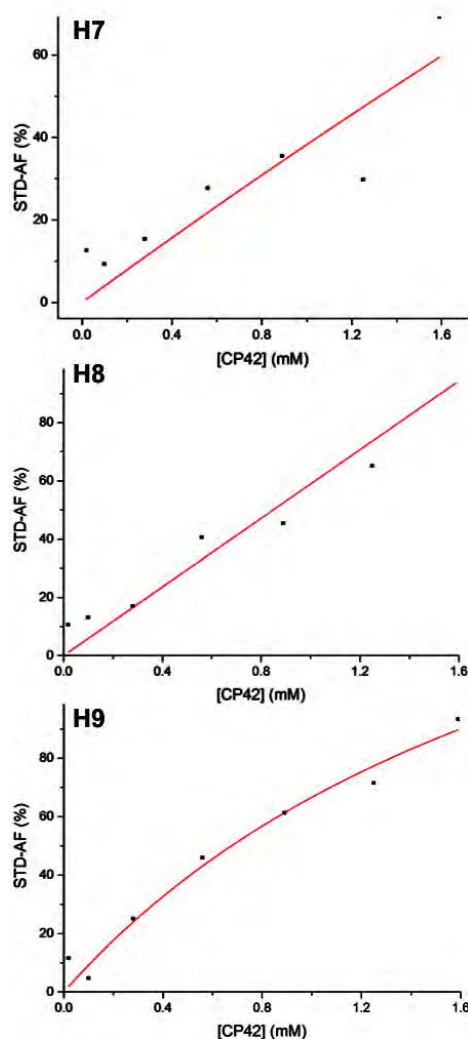


Figure 4.13: Binding isotherms fits for CP42. The STD-AF growth curve of each CP42 protons exhibiting saturation (dots), and the fit to Equation (4.2) (red line).

The apparent equilibrium constants (K_D) of each protons that yielded STD-NMR signals are listed in Table 4.1. All individual protons from the aromatic region had high affinity to CPG2; however, the magnitude of the affinities observed was different. The glutamate group in the "warhead" was in close proximity to CPG2, but displayed very weak affinity to CPG2 suggesting the importance of it for CPG2 recognition. The equivalent galactosylated ligand displayed very different tendencies. Indeed, CP67 glutamate protons showed very high affinity to CPG2 (ranging from 75 to 174 μM), whereas CP42 displayed very weak binding. The binding affinities of the same ligands determined by the octet RED system by Mologic were universally stronger (smaller K_D) (Figures 4.1 and 4.1). The octet RED system (ForteBio, USA) is an instrument that specialises in measuring the affinity of small molecule binding to a therapeutic target. The purified CPG2 was immobilised onto an amine reactive biosensor and assayed against the inhibitors of interest. The lower K_D values could be attributed to the presence of a protein layer on the biosensor and the use of bovine serum albumin (BSA) as a blocking agent to prevent non-specific binding events. Meanwhile, the higher K_D values observed with STD-NMR are possibly due to NMR methods giving an averaged picture of the ligand in its bound and free states (Fielding et al. 2005).

Table 4.1: Apparent binding constants. The K_D values obtained from each individual curves fits, presented in Figures 4.11, 4.12 and 4.13, are unveiled next to their respective protons. The mean value was calculated as $437 \mu\text{M}$, $227 \mu\text{M}$ and in the millimolar range for CP6, CP67 and CP42, respectively. The standard deviations are $374 \mu\text{M}$ and $94 \mu\text{M}$, for CP6 and CP67 respectively.

Ligand	Proton number	K_D (mM)
CP06	H1	0.339
	H2	0.450
	H3	1.326
	H4	0.456
	H5	0.234
	H6	0.325
	H7	0.199
	H8	very high
	H9	very high
CP67	H1	0.330
	H2	0.254
	H3	0.289
	H4	0.251
	H5	0.346
	H6	0.333
	H7	0.140
	H8	0.174
	H9	0.107
	H9'	0.075
	H10	0.203
CP42	H7	24.674
	H8	very high
	H9	2.277

4.4 Protein-detected NMR experiments.

Most of the protein-based methods rely on chemical shift perturbations as barometer for intermolecular binding, monitoring the changes in the NMR spectrum of the protein upon ligand binding. This technique was first applied to study structure–activity relationships obtained from NMR by Shuker et al. (1996)). Drug-design has known exceptional progress since the likelihood of identifying and characterising potential weak binding molecules at the surface of drug-targets using heteronuclear NMR experiments.

4.4.1 ^1H - ^{15}N -HSQC of Free CPG2.

Typically ligand binding studies involve uniform isotopic labelling of the protein as protein signals are monitored using heteronuclear NMR experiments. To overcome problems associated with higher molecular weight proteins, isotope labelling and deuteration were used.

Here, we managed to remove protons to 45-50% in a random fashion in order to reduce external relaxation contributions. 226 μM of his-tagged CPG2 (approximately 4 mg per NMR sample) in 10 mM phosphate buffer at pH 8 was used for NMR analyses. ^1H - ^{15}N -HSQC is the most common experiment and displays all ^1H - ^{15}N correlations. This spectrum is a fingerprint of the folded protein and is usually the first heteronuclear experiment carried out on proteins. From it, it could be judged whether other experiments are likely to succeed because each protein has a unique pattern of N-H correlations for each residues

(see Section 1.6.2 in Chapter 1). As so, it can also be used as a tool to optimise the NMR sample conditions (temperature, pH, etc). Spectra were acquired at three different temperatures: 25°C (with 64 scans), 30°C (128 scans) and 37°C (256 scans) (Figure 4.14). It was assumed that increasing the temperature would reduce the rotational correlation time and improve the spectrum (Figure 4.14-B and C). Unfortunately, the protein started to precipitate at 30°C and 37°C. It was decided that 25°C was the optimal temperature at which to run NMR experiments. Secondly, the dispersion of signals in the spectrum A of the protein is far beyond the envelope of signals seen in the spectrum B, where clustered peaks are observed around 7.5 and 9 ppm. This clearly reflects that nuclei in spectrum A are in many distinct environments and suggests proper folding of the protein. 236 of the 420 possible peaks were observable in this sample (e.g. 60%, Figure 4.14-A). While we had successfully expressed CPG2₂₃₋₄₁₅ in a soluble, active form in *E. coli*, the resulting protein displayed limited long-term stability (activity deteriorated 24 hours post-purification) and a propensity to aggregate over time in Tris buffer containing 50 mM NaCl (pH 7.2) as seen in Chapter 3.

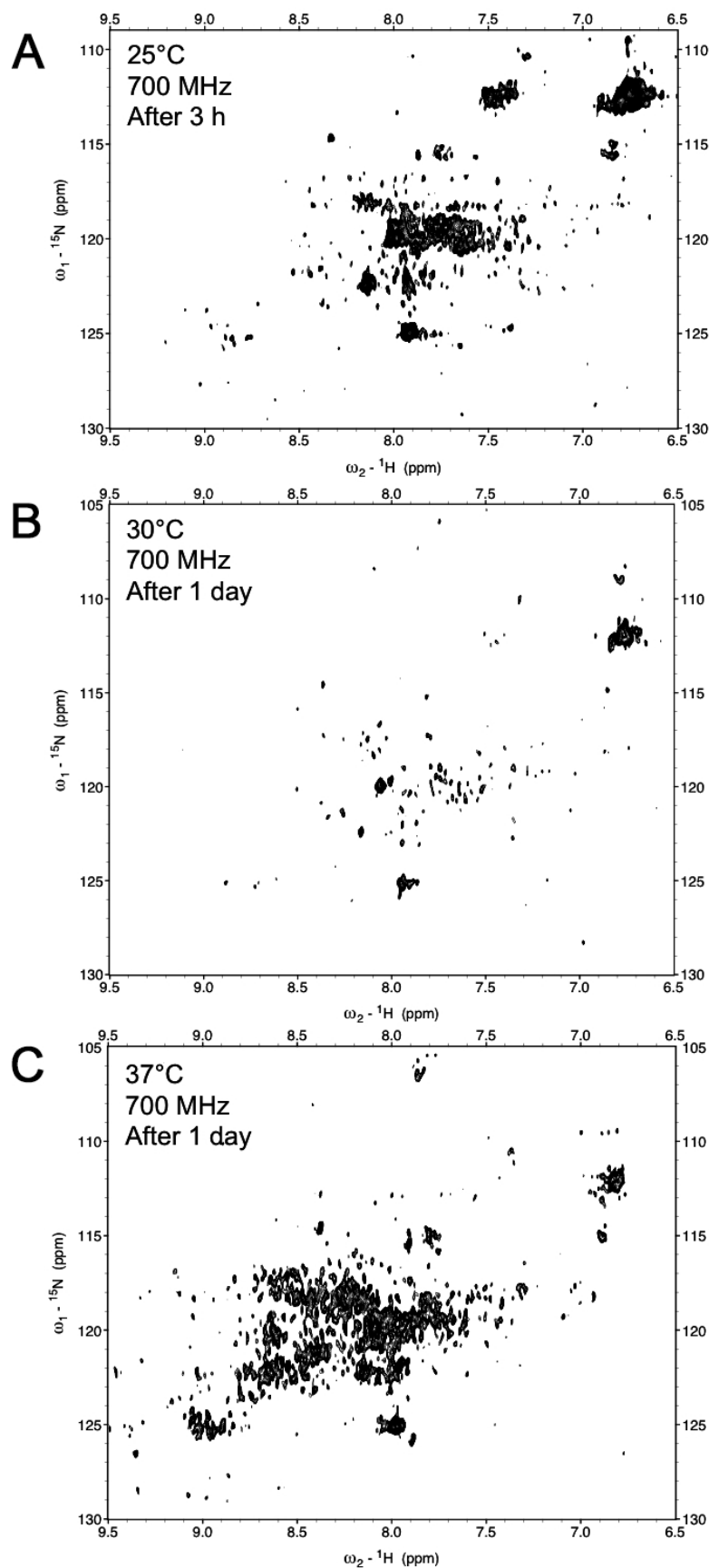


Figure 4.14: Three CPG2 ${}^1\text{H}$ - ${}^{15}\text{N}$ -HSQCs at different temperatures and number of scans. A) CPG2 ${}^1\text{H}$ - ${}^{15}\text{N}$ -HSQC (4 mg/180 μL) at 25°C, 3 hours. B) CPG2 ${}^1\text{H}$ - ${}^{15}\text{N}$ -HSQC (4 mg/180 μL) at 30°C, 1 day. C) CPG2 ${}^1\text{H}$ - ${}^{15}\text{N}$ -HSQC (4 mg/180 μL) at 37°C, 1 day (precipitation).

4.4.2 TROSY-NMR Characterisation of 42 kDa CPG2₂₃₋₄₁₅ at Two Fields and Two Temperatures.

After successful production of high-yields of soluble, active and isotopically labelled CPG2₂₃₋₄₁₅ that was stable at 18°C for at least five days as described in Chapter 3, we wished to evaluate the potential of solution NMR spectroscopy to study the structure, substrate / ligand-binding, catalytic mechanism and dynamics of this protein in solution. NMR has not been used to study CPG2 previously, and allows study of this enzyme in solution without the need to generate crystals. The molecular weight of CPG2₂₃₋₄₁₅ (42 kDa) and its dynamic nature (L-shaped protein) makes NMR assignment and subsequent structural investigation of the full-length enzyme an ambitious goal. Thus we applied TROSY (transverse relaxation-optimised spectroscopy) based techniques to measurements of ²H/¹⁵N/¹³C-labelled CPG2₂₃₋₄₁₅. TROSY, pioneered by Wüthrich, decreases the fast T₂-mediated decay of amide-proton NMR signals, by exploiting constructive use of the two effects that determine transverse relaxation in large proteins: dipole-dipole coupling (*DD*) between two spins and chemical shift anisotropy (*CSA*) of the two spins (Pervushin et al. 1997; Zhu et al. 1999). This relaxation interference minimises line-broadening due to limited molecular tumbling in large proteins and can result in significant improvements in resolution and sensitivity (Ollerenshaw et al. 2003; Wuthrich 1998). To further minimise loss of sensitivity due to increased sample conductivity when using the cryogenic NMR probe, data were then acquired for CPG2₂₃₋₄₁₅ dissolved in the low-conductance buffer solution (20 mM Tris, pH

7.3 and 50 mM L-arginine, 50 mM L-glutamate) after cleavage of the his-tag.

^1H - ^{15}N -TROSY-HSQC spectra were measured at two field strengths (700 and 900 MHz²) and two temperatures (25°C and 37°C) and the results are shown in Figure 4.15. At both temperatures, we observed the expected improvement in linewidths at 900 MHz when using TROSY-based methods (the TROSY effect has been shown to yield the narrowest line widths at 900 MHz), but a substantial reduction in the number of peaks observed as annotated in Figure 4.15 (Keeler 2011). At both field strengths, increasing the temperature to 37°C yielded a noticeable improvement in peak width and spectral overlap (Figures 4.15-B, D) and a significant increase in the number of peaks observed. This was unsurprising since peak width, and thus the resolution and signal-to-noise ratio, can be significantly improved by varying the acquisition temperature, which affects the tumbling time and relaxation properties of large proteins (Keeler 2011). However, sustained exposure to 37°C temperatures lead to aggregation of the enzyme after two days in the low-conductance buffer solution. Acquisition at 25°C (700 MHz) reduced aggregation, and yielded a spectrum containing 318 of the 420 theoretical HSQC peaks (or 76%) expected from the primary sequence of CPG2₂₃₋₄₁₅ (Figure 4.15-A).

²Dr Sara Whittaker (School of Cancer Sciences, University of Birmingham) collected the NMR data using the 900 MHz Bruker Instrument.

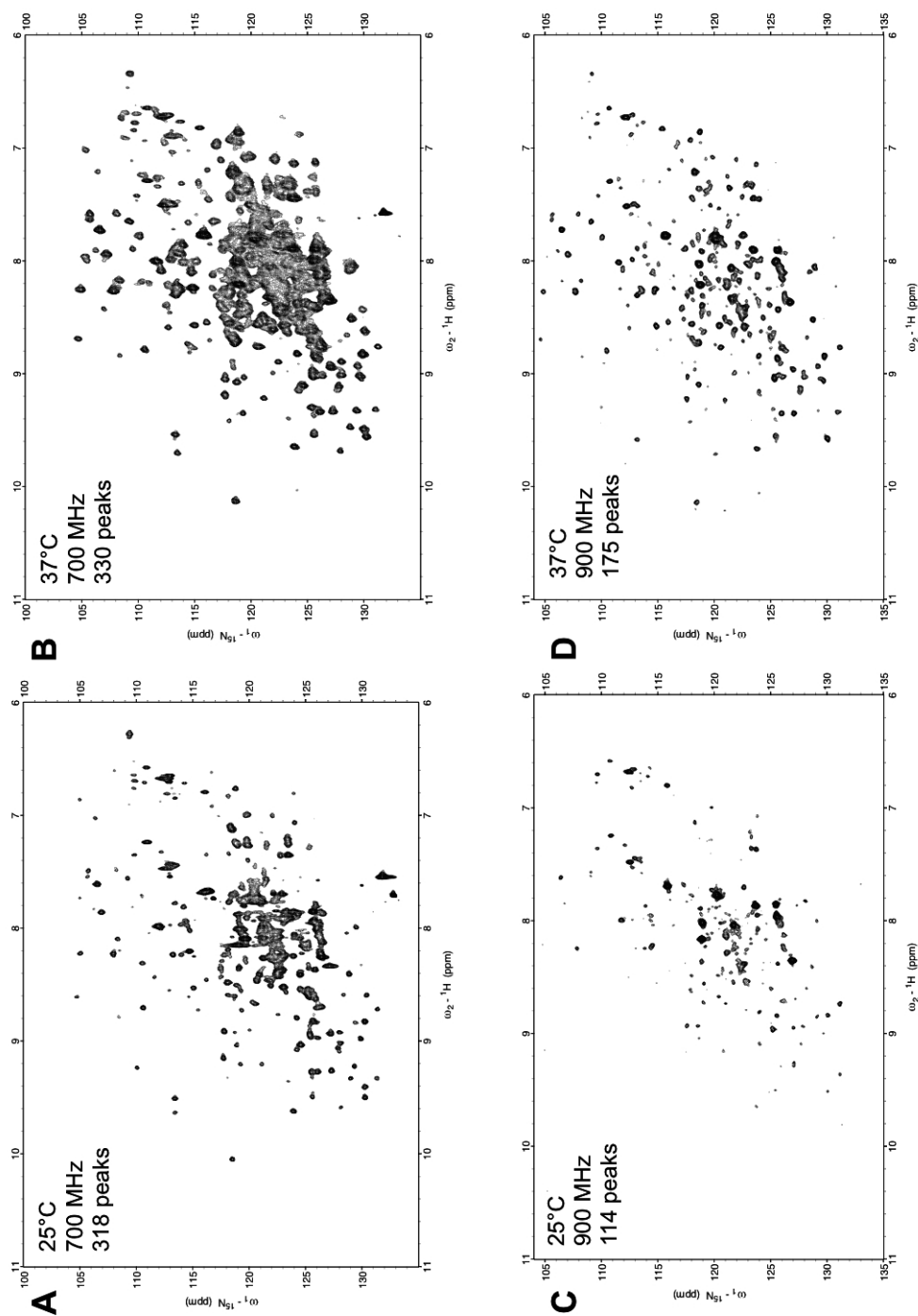


Figure 4.15: ^1H - ^{15}N TROSY-HSQC spectra acquired at different fields and temperature. ^1H - ^{15}N TROSY-HSQC spectra acquired for ^1H , ^{15}N , ^{13}C -labelled CPG₂₃₋₄₁₅ solubilized in 20 mM Tris (pH 7.3) containing 50 mM L-arginine and 50 mM L-glutamate acquired at A) 700 MHz / 25°C, B) 700 MHz / 37°C, C) 900 MHz / 25°C and D) 900 MHz / 37°C. The number of observed peaks (out of a possible 420 predicted from sequence) are listed.

Using these optimised parameters, we acquired a ^1H - ^{15}N TROSY-HSQC spectrum of $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ -labelled CPG2₂₃₋₄₁₅ solubilised in 20 mM Tris buffer (pH 7.3) containing 100 mM NaCl (Figure 4.16). Although this buffer contained NaCl concentrations that may lead to sensitivity loss when using the cryogenic NMR probe, it yielded protein samples with the highest stability while retaining activity. No loss of sensitivity was observed in this buffer at 25°C when compared to the equivalent spectrum acquired in buffer containing 50 mM L-arginine, 50 mM L-glutamate (Figure 4.15-A), suggesting that the cryoprobe was not negatively impacted by the presence of 100 mM NaCl (Kelly et al. 2002). Quite the opposite was observed as this spectrum displayed improved signal to noise, narrower linewidths, and an increase in the number of observable peaks (342 peaks, or 81% of the expected peaks) as compared to the equivalent spectrum acquired in the Arg/Glu-containing buffer.

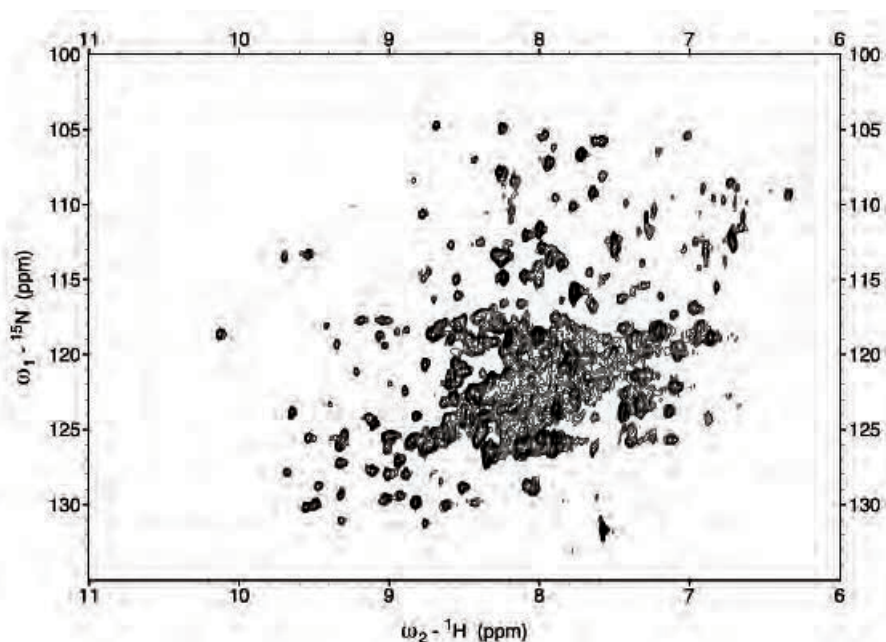


Figure 4.16: ^1H - ^{15}N TROSY-HSQC spectrum of $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ -labelled CPG2₂₃₋₄₁₅. ^1H - ^{15}N TROSY-HSQC spectrum acquired for $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ -labelled CPG2₂₃₋₄₁₅ solubilised in 20 mM Tris (pH 7.3) containing 100 mM NaCl. The spectrum was acquired at 25°C and 700 MHz, and yielded 353 resolved peaks.

4.4.3 Unravelling Binding by Using Simple HSQC NMR Experiments.

In these protein-detected experiments, ligand binding is observed via chemical shift changes in the ^1H - ^{15}N HSQC spectrum of the protein induced by the binding of the ligand. In this section, we decided to focus on the first set of inhibitors CP06 and CP67, the warhead and its galactosylated alternative that showed inhibitory activity. Despite the large number of expected resonances and some overcrowded regions in the spectrum, a large number of peaks were resolved, thereby producing a large number of potential probes within the protein upon addition of CP06 and CP67. To our knowledge, the NMR assignment of CPG2 have not been elucidated, thus preventing the localisation of the amide resonances in the CPG2 3D X-ray structure. However, we expected these correlation peaks to provide a sufficiently large illustration, enabling characterisation of the kinetic events taking place at the protein surfaces in the presence of these two ligands by titrating increasing concentration of ligand into the CPG2 solution (see Section 1.6.2 in Chapter 1).

Several interaction experiments with different ligand concentrations were performed to obtain a K_D . A mixture of approximately 0.45 mM CPG2 protein sample with increasing CP06 ligand concentration samples were prepared to probe the binding events (Figure 4.17).

1[h!]

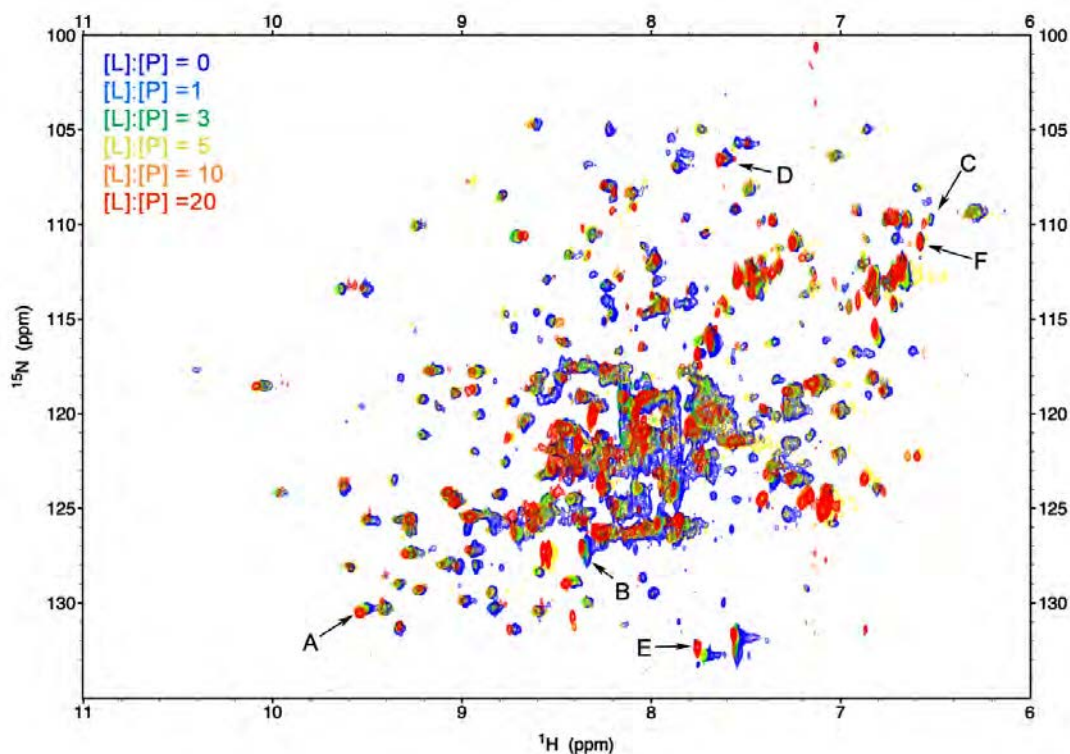


Figure 4.17: ^1H - ^{15}N -HSQC spectra of CPG2 acquired in the presence of CP6 ligand. Titration were carried out with 1:1, 3:1, 5:1, 10:1 and 20:1 ligand to protein concentration ratio. The spectra shows changes in the protein signals upon additions of the CP06.

Most of the ^1H - ^{15}N peaks in Figure 4.17 were not affected by the presence of ligand (peak F), and only a small set of isolated peaks moved (peaks A to E). Their behaviour were thoughtfully investigated in Figure 4.18. To start with, the ligand binding kinetics are likely in the fast exchange regime on the NMR chemical shift timescale (Williamson 2013). All the peaks moved smoothly from the free protein chemical shift (blue) to the chemical shift of the bound protein (red). However, different behaviours in chemical shift changes were observed. Chemical shifts of peak A and C displayed the same shape throughout ligand titration suggesting that they were in fast exchange rate; chemical shifts of peaks D and E became broader upon ligand titration, but sharpened up again close

to saturation, suggesting they were at the limit of the very fast exchange rate (Figure 4.18). Chemical shift changes for peaks B and E displayed a linear shift in the ^1H - ^{15}N correlation peak upon titration, while chemical shift changes for peaks D and C were not linear which was indicative of an intermediate formed during the titration demonstrating a conformational change in the bound protein (Figure 4.18-D and C).

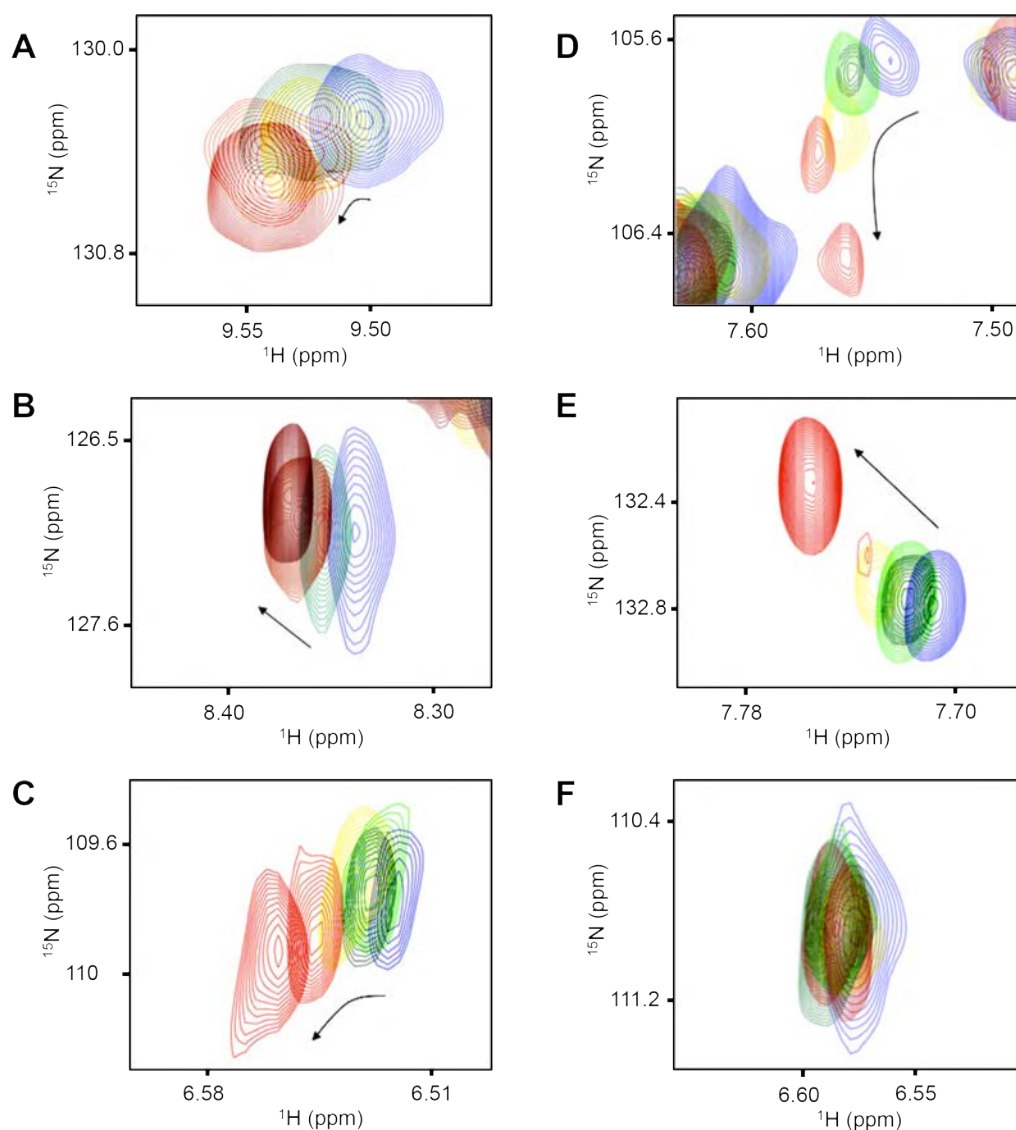


Figure 4.18: Close-up of ^1H - ^{15}N HSQC spectra of CPG2 acquired in the presence of CP06 ligand. Titration were carried out with 1:1, 3:1, 5:1, 10:1 and 20:1 ligand to protein concentration ratio. The figure shows some of the significant chemical shift changes upon additions of the CP06.

Multiple binding curves were determined by titrating each ligand into separate samples of $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ -labelled CPG2 and acquiring ^1H - ^{15}N HSQC spectra at as many as 5 different ligand to protein concentration ratios ranging from 1:1 - 20:1. Chemical shift changes in both ^1H and ^{15}N dimensions were logged and fitted by non-linear regression analysis to Equation (4.3) (Fielding 2007; Sudmeier et al. 1980).

$$\Delta_{\text{obs}} = \Delta_{\text{max}} = \frac{(K_{\text{D}} + [L]_0 + [P]_0) - \sqrt{(K_{\text{D}} + [L]_0 + [P]_0)^2 - (4[P]_0[L]_0)}}{2[P]_0} \quad (4.3)$$

The measured K_{D} values are shown in Figure 4.19. These HSQC titration derived K_{D} (41 to 493 μM) compared favourably with that derived from the STD-NMR titration experiments (169 to 1326 μM) and octet RED system (88.5 μM).

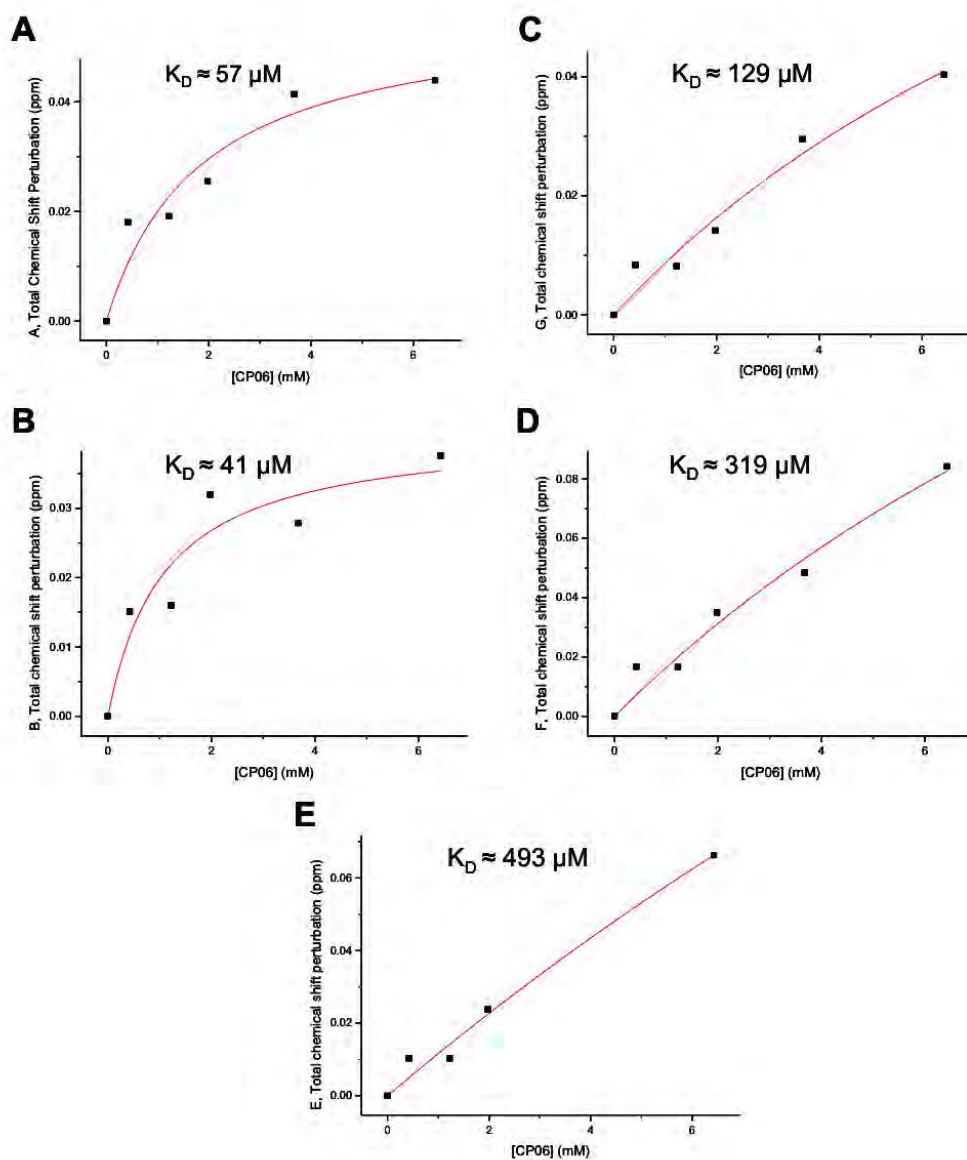


Figure 4.19: CP06 curve fits. Fits for all significant chemical shift differences using Equation (4.3) are shown. The K_D value obtained from each individual curve fit is displayed next to the respective curve.

HSQC titration experiments were also repeated with CP67 as a ligand, the galactosylated derivative of the CP06 inhibitor, as shown in Figure 4.20. Upon addition of CP67, changes in ^1H - ^{15}N chemical shift correlations were observed indicating local changes in environment caused by CPG2-CP67 complex formation. Some peaks were still visible and displayed no chemical shift changes (Figure 4.20-A,B and C), indicating that these resonances were not affected by CP67 binding. This implied that these residues may be located further away from the CP67 binding site, as there were only limited chemical shift changes. NMR spectra also indicated small chemical shift changes (Figure 4.20-D). Additionally, a great number of peaks disappeared as soon as the CP67 was titrated in (Figure 4.20-E) and other peaks started to shift (Figure 4.20-F). Some NMR peaks probably broaden beyond detection by binding to CP67, which may quench internal motion of the protein. But the great number of anomalous peak broadening upon titration suggested that these protein residues may have been directly involved in CP67 recognition and binding. Moreover, binding of CP67 to CPG2 dramatically changed the global molecular tumbling motion of the protein due to the lack of any visible narrow NMR resonances.

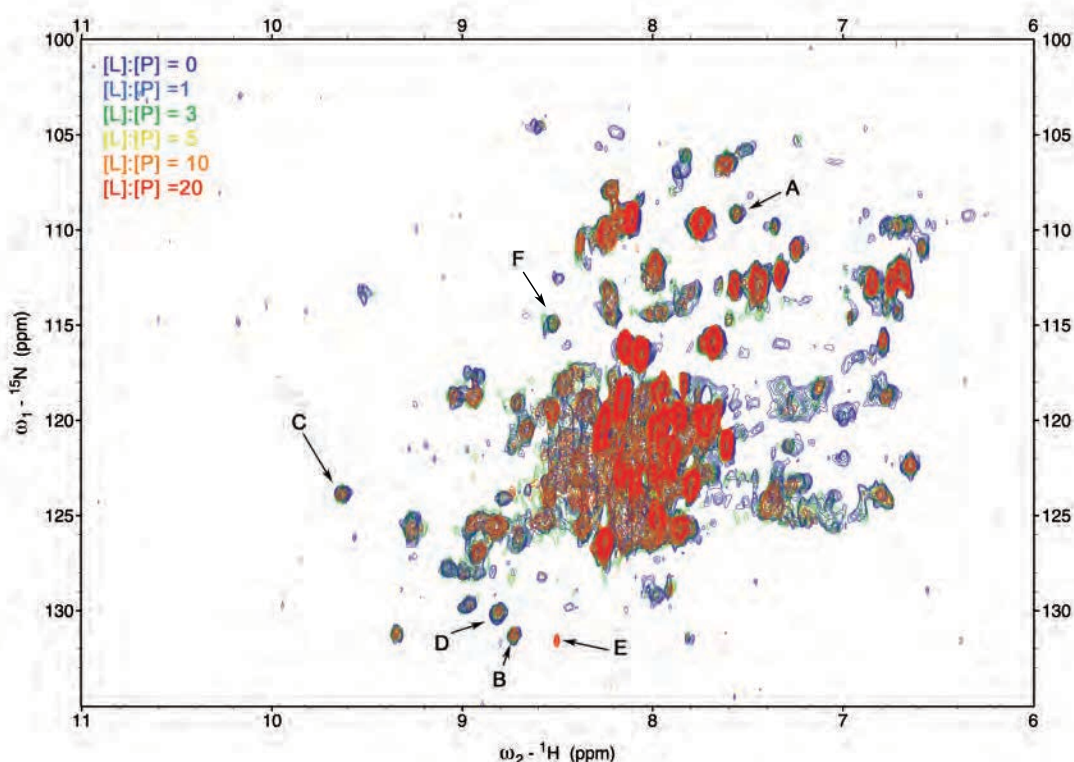


Figure 4.20: ^1H - ^{15}N HSQC spectra of CPG2 acquired in the presence of CP67 ligand. Titrations were carried out with 1:1, 3:1, 5:1, 10:1 and 20:1 ligand to protein concentration ratio. The spectra shows some changes (i.e broadening) in the protein signals upon additions of the CP67.

Other ^1H - ^{15}N correlation appear at higher CP67 concentrations (Figure 4.21-E). This indicates that in the presence of the galactosylated tail, this protein residue was in slow exchange on the chemical exchange rate as the free signal does gradually decrease in intensity as the bound signal increases after addition of an excess of CP67 ligand as shown in Figure 4.21 (Williamson 2013). On the other hand, Peak F in Figure 4.21 displayed an intermediate motion because during the titration the NMR signal for this residue seemed to broaden and form a doublet, but as soon as the protein was saturated with CP67, the peak disappeared. Because of signals broadening during ligand titration, it was not possible to extract the K_D for CP67.

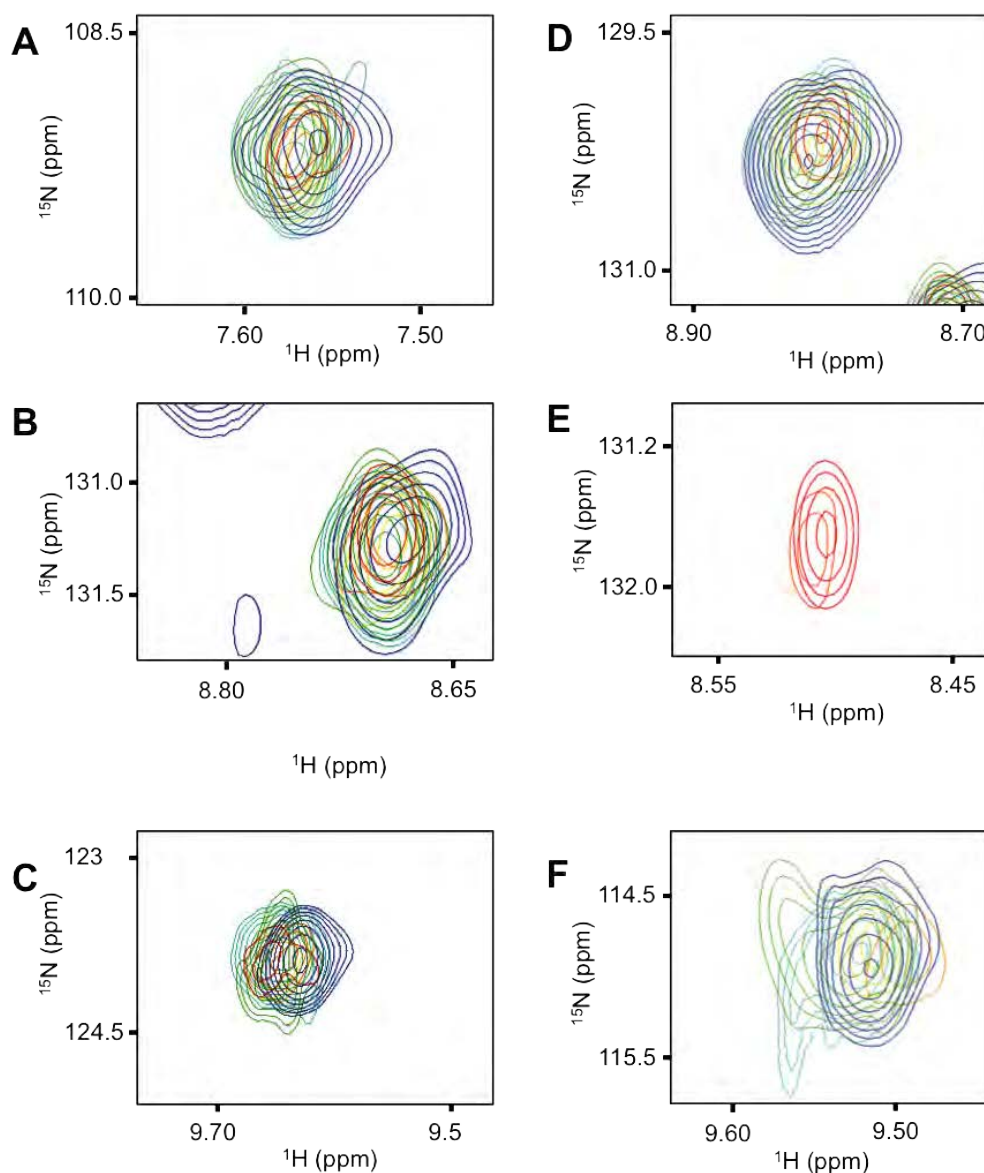


Figure 4.21: Close-up of ^1H - ^{15}N HSQC spectra of CPG2 acquired in the presence of CP67 ligand. Titration were carried out with 1:1, 3:1, 5:1, 10:1 and 20:1 ligand to protein concentration ratio. The figure shows some of the significant chemical shift changes upon additions of the CP67.

4.5 Conclusion

This chapter describes how relevant ligand-protein interactions can be studied without chemical shift assignment of the 2D ^1H - ^{15}N HSQC spectrum of CPG2.

The effects resulting from the binding of a ligand can be categorised into two groups: the global effects and the local effects (Heller and Kessler 2001). The former are size-dependant and therefore are suited to the observation of the ligand. In general, upon binding to a macromolecule, the apparent molecular weight will change dramatically the size by several orders of magnitude. Those changes can be studied using NMR experiments, such as the DOSY experiment (Goldflam et al. 2012). It will reveal binding, but cannot provide structural information about the inhibitors or the binding process. On the other hand, the global effects of binding that are restricted to the binding region can be monitored on both CPG2 and the inhibitors, using STD-NMR and chemical shift perturbation methods (Heller and Kessler 2001).

The inhibitors provided and produced by Mologic, shown in bold in Figures 4.1 and 4.2, were used in a ligand-based approach via the STD-NMR experiment. The initial idea was to investigate the reason why CP06 and CP39 upon galactosylation, retained or lost inhibition strength. Unfortunately, 1D ^1H NMR of CP39 showed degradation. Hence, we were left with inhibitors CP06, CP67 and CP42 for further analysis. Analysis of the three inhibitors showed that the glutamate moiety is indeed essential for CPG2 recognition and / or binding, as protons of the glutamate group in CP06 (Figure 4.7) and CP67 (Figure 4.8) showed saturation, while none of the glutamate protons in CP42 seemed to be in close proximity to CPG2 (Figure 4.9). However, the CP06 and CP67 glutamate H7, H8, H9/9' protons that were revealed to be close to CPG2 did not show very high binding constants (Table 4.1). H7, H8 and H9 protons of CP42 (found in the benzyl group, between the galactosylated and glutamate

groups) were the only protons to display any closeness to CPG2 (Figure 4.9), but K_D values for these protons were very high suggesting very weak binding to CPG2 in accordance with the RED octet values (Figure 4.2). The change in the aromatic ring type seemed to have an effect on the inhibitory effect of these ligands. Indeed, CP06 and CP67 inhibitors revealed that the naphthalene ring present in these inhibitors may play a significant role in molecular recognition by CPG2 (Figures 4.7 and 4.8). The protons of the naphthalene ring showed very micromolar range K_D on average below 0.5 mM for CP06 and below 0.35 mM for CP67, except for proton for CPO6-H6 proton (Table 4.1). CP06-H6 proton displayed an unusually high binding K_D constant in the millimolar range compared to its neighbouring protons. This is probably due to tight binding of the ring to CPG2 affecting the final K_D . The interesting aspect involving CP67 analysis was that none of the protons of the galactosylated moiety revealed proximity to CPG2 protein. This results indicated that, firstly, the galactosylated moiety was available for the hepatocytes receptors to bind to for blood clearance, and secondly, the galactosylated moiety did not interfere with the rest of the CP67 compound involved in CPG2 recognition and / or binding. As a result, CP06 and CP67 were taken forward for chemical shift perturbation investigation using ^1H - ^{15}N HSQC experiments titration.

Usually protein-based method is used to enable to resolve the ligand binding footsteps on the target protein. However, NMR structural studies of large proteins such as CPG2 face distinct challenges as it suffers from line-broadening, frequency degeneracy (i.e. resonances that have overlapping frequencies) and spectral crowding, which is confinement of a large number of

NMR signals within a given spectral region due to a large number of residues as we observe in Figure 4.14. For optimal data collection, adequate perdeuteration (Chapter 3) and TROSY was used (Pervushin et al. 1998). By combining a suitable labelling strategy with the appropriate pulse sequence, we were able to generate an acceptable ^1H - ^{15}N TROSY-HSQC spectrum. Before assigning the signals observed the NMR spectra to the nuclei in CPG2, we decided to use a very simple method to study binding to protein: chemical shift perturbation. In addition to exposing potential binding sites, it can present detailed affinity information.

^1H - ^{15}N TROSY-HSQC were acquired as increased concentration of CP06 and CP67 respectively were titrated in CPG2. The two ligand did not behave similarly upon titration. CP06 displayed linear chemical shift perturbation and CP67 exhibited large line-broadening as ligand concentration was increased. CP67 displayed very fast exchange rate on the NMR time scale as resonances were averaged for the free and bound species of CPG2 (Figure 4.18) (Bryant 1983). In some cases, shifts were not in a straight line, which was indicative of an intermediate conformation formed during the titration displaying a conformational transition in the bound protein (Williamson 2013). Fitting the chemical shift perturbations to a 1:1 binding model using Equation (4.3) produced the best fits to the data and allowed calculation of the binding K_D constants for the few residues that suggested interaction between CPG2 and CP06. All K_D values yielded to micromolar range values between 40-493 μM (Figure 4.19). These values were very similar to the binding constants observed with STD-NMR for CP06, where K_D were ranging below 0.5 mM.

Unfortunately, considering the pairing of line broadening with spectral crowding upon titration as observed in Figure 4.20, K_D could not be obtained for CP67 even when it was showing chemical shifts (Figure 4.21-F).

To conclude, it is evident from the STD-NMR data and chemical shift perturbation data for CP06 and CP67 that residues are affected by the inhibitors binding that could be explained by direct interaction between CPG2 and the different structural parts of the inhibitors identified via STD-NMR experiments: the aromatic ring and the glutamate moiety. However, further work is required to confirm the scope of these ligand-induced effects and techniques. In the future, the determination of the solution structure of CPG2 could further characterise the binding sites, and conformational effects of these inhibitors in order to design more potent drugs.

Chapter 5

Site-directed Mutagenesis of Positions Within or Close to the Regions of CPG2 Directly Related to Substrate Binding

5.1 Introduction.

The X-ray crystallographic structure of CPG2 has been determined (Rowse et al. 1997). It contains a co-catalytic domain and a dimerisation domain. Analysis of the co-catalytic Zn(II) active site of CPG2 reveals a distinctive similarity to the active site of *Aeromonas proteolytica* aminopeptidase AAP. CPG2 contains two metal ions bound in its active site with a (μ -aqua)(μ -carboxylato) dizinc(II) core as shown in Figure 5.1 (Remaut et al.

2001). The coordination geometry of each Zn(II) ion is tetrahedral, with Zn1 being liganded by a carboxylate oxygen of Glu176 and an imidazole nitrogen of His385. Similarly, Zn2 is coordinated by α -carboxylate oxygen of Glu200 and imidazole nitrogen His112. These residues at the active site co-ordinates each of the metal atoms separately and allow conservation of a Zn1 to Zn2 distance of 3.3 Å. The Asp141 residue co-ordinates both zinc ions at once while, near the co-catalytic Zn(II) active site of CPG2 residues, a glutamate residue (Glu175) forms a hydrogen bond to the zinc bridging water molecule (Rowsell et al. 1997). Glu175 probably functions as a general acid / base during catalysis by making the water molecule reactive, in a similar manner to Glu151 in AAP (Bzymek and Holz 2004).

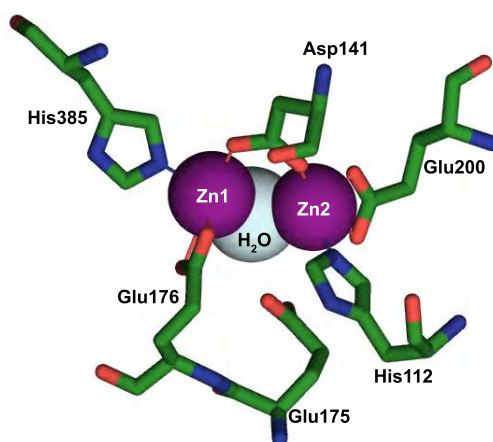


Figure 5.1: Zinc binding domain. View of CPG2 catalytic site based on the X-ray crystallographic structure by rowsell1997 (PDB: 1CG2). The site is shown in a *ball and stick* representation. Zinc ions are represented as *purple spheres* and the water molecule as a *grey spheres*. The residues involved in catalysis are also displayed. [Pymol Viewer]

In ADEPT, the pro-drug is converted from an inactive form to a cytotoxic agent by CPG2. As the mechanism of action of CPG2 has not been elucidated yet, the spectrum of inhibitors which are used in medicine is limited. The usual features that seem to be crucial include an α -carboxylated moiety on the L-

glutamate residue and a benzyl ring close to the carbonyl carbon of the amide bond (Khan et al. 1999; DeAngelis et al. 1996). Thiolate-containing inhibitors based on a thiocarbamate moiety attached to a benzene ring have also been reported for CPG2 but were only moderately effective competitive inhibitors and displayed toxicity towards human colonic epithelial cancer cell lines (LS174T cells), possibly allowing their use as inhibitors in ADEPT (Khan et al. 1999). As described in Chapter 4, Mologic has since developed peptide-based ligands that targeted the CPG2 enzyme. These peptides must also carry sugar residues that enable them to be bound by the asialoglycoprotein receptor of hepatocytes, to ensure efficient clearance of the inhibited CPG2 complex via the liver. As shown previously in Chapter 4, these inhibitors are made of a warhead, which consists of an aromatic group and a glutamate moiety, attached to a galactosylated tail (see Figures 4.1 and 4.2 on pages 112 and 113 respectively). The warhead is directly implicated in CPG2 recognition and its inhibition, whereas the galactosylated tail is present to ensure rapid removal of the inhibited CPG2 complex from the bloodstream.

The overall structure of CPG2 is very similar to the structures of corresponding domains in the related enzymes aminoacylpeptidase (AAP) and membrane-bound human glutamate carboxypeptidase II (GCPII) (Lindner et al. 2003). Their mechanisms of catalysis were proposed via the use of computational methods (Klusák et al. 2009; Schürer et al. 2004). Using information extracted from these studies, a binding mode hypothesis was established for the CPG2-MTX complex by applying molecular docking and molecular dynamics simulation (Turra et al. 2012). Turra et al. (2012)

established that MTX was interacting at a very specific position in the CPG2 catalytic region, which may direct the substrate recognition and cleavage mechanism (Figure 5.2) (Turra et al. 2012). CPG2 seemed to accommodate MTX in the active site by adjusting the N-terminal pteroate moiety of MTX in an adjacent pocket (P1). The pteridine group looked to remain secure via hydrogen bonding interaction involving nitrogen atoms from pteridine and the carbonyl oxygen from Ser210, Lys208 and Thr357. The glutamate side-chain pointed towards the protein surface and appeared to be hydrogen bonded to Arg324 and Gly360 residue, which are located in another pocket adjacent to the active site (P2). Rowsell et al. (1997) showed that the Arg324 mutation to alanine resulted in a mutant CPG2 with very low activity towards methotrexate. Arg324 is obviously playing an important part in catalysis but its exact role still remains unknown. None of the other residues thought to be involved in substrate recognition was studied by site-directed mutagenesis yet.

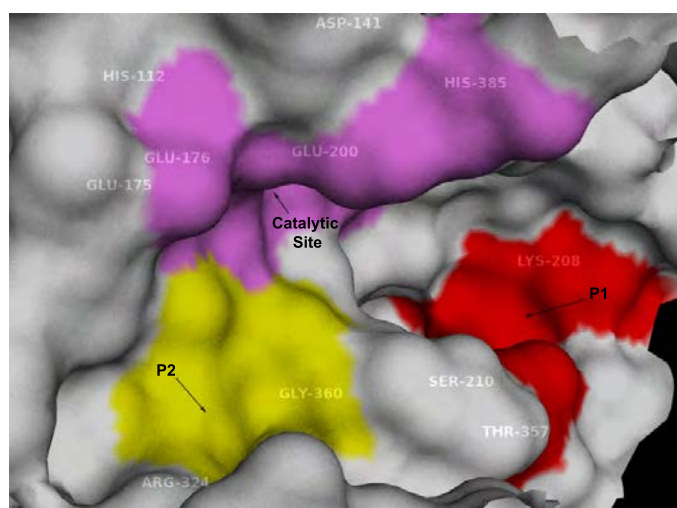


Figure 5.2: Two possible pockets in the CPG2 active site. The protein external surface of interior cavities and pockets in the active site are shown based on the X-ray crystallographic structure by Rowsell et al. (1997) (PDB: 1CG2). The catalytic site's molecular surface is coloured in pink and the two possible P1 and P2 pockets are displayed in red and yellow, respectively. The residues possibly involved in substrate molecular recognition are also shown. [Pymol Viewer]

Despite the above-stated efforts, the reaction mechanism of CPG2 is not fully understood and, to date, no co-crystallised complex has been published. This work sought to further complement the structural results obtained from CPG2 X-ray crystal structure, and computational simulations of CPG2-MTX complex, using site-directed mutagenesis studies to focus on the effects of more comprehensive residue mutation in both P1 and P2 pockets, and also the catalytic site of CPG2, against MTX and competitive inhibitors produced by Mologic (see Figures 4.1 and 4.2 on pages 112 and 113) using the well-established CPG2 spectrophotometric assay (Bhatia et al. 2000).

5.2 Site-Directed Mutagenesis, Expression and Purification of CPG2 Mutants.

Mutagenesis was used to report the site of functional importance for MTX (and its competitive inhibitors) binding. This study was aimed at describing the contribution of the different residues introduced above, and previously proposed in theoretical studies to MTX substrate recognition and / or binding. Figure 5.3 shows the different residues possibly participating in substrate recognition and catalysis. The intention was to mutate all the residues shown in Figure 5.3 to alanine, a small uncharged amino acid, to establish whether the side-chains of these various residues played a critical role in CPG2 bioactivity. The residues taken forward for the study were H112, H385, E175, E176, E200, D141, K208, S210, T357 and G360.

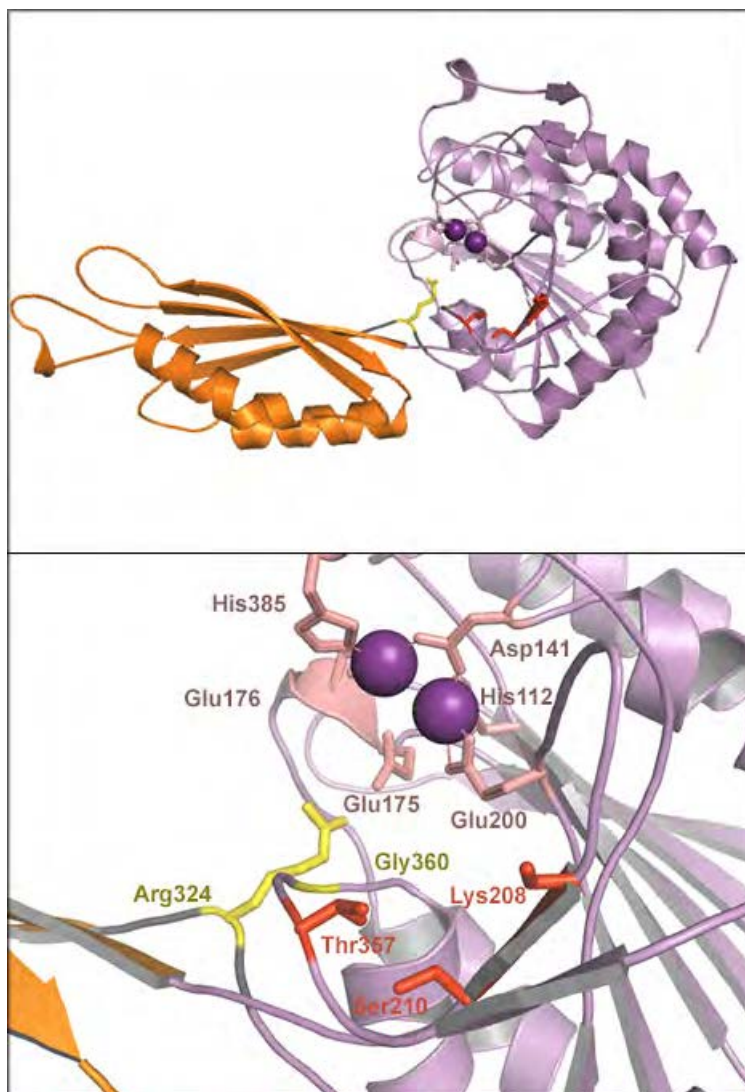


Figure 5.3: Residues mutated to alanine in the active site. *Top panel* shows the different residues within CPG2 proposed to make up binding sitemapped onto the x-ray crystallographic structure by Rowsell et al. (1997) (PDB: 1CG2). *Bottom panel* is a close-up of the region possibly involved in MTX catalysis. In pink are shown the residues involved in catalysis, in red the residues possibly interacting with the pteridine moiety of MTX and in yellow the residue that may be in contact with the glutamate group of MTX. [Pymol Viewer]

Site-directed mutagenesis¹ were performed using the QuickChange site-directed mutagenesis kit (Stratagene, La Jolia, CA) according to manufacturers instruction. The oligonucleotide primers used to introduce the alanine mutation at each residue position into the pet151-CPG2₂₃₋₄₁₅ plasmid are shown in Table 2.3 on page 37. The presence of the mutation was verified by DNA sequencing (GATC, UK). Successful mutations were carried out for H112A, H385A, E175A, E200A, D141A, K208A, S210A and T357A. E176A and G380A mutations were not obtained, as transformation of plasmids pet151-E176A and pet151-G380A into *E. coli* were not effective.

All mutant constructs were expressed and purified using the same purification protocol used to produce unlabelled CPG2₂₃₋₄₁₅ in order to avoid additional optimisation steps for purification (Section 3.3). The mutants were stored in 20 mM Tris (pH 7.3) containing 100 mM NaCl and 20% glycerol, at -20°C until future use. Figure 5.4 shows typical purification steps and characterisation by MALDI-TOF carried out for each mutant. The second round of nickel affinity chromatography (Figure 5.4-A, lane FT) produced a very pure mutant protein sample, as confirmed by MALDI-TOF (Figure 5.4-B) for T357A. This removed the need for an extra purification step using size-exclusion chromatography. MALDI-TOF of the T357A mutant showed a monomeric peak at 41,683.91 Da (theoretical mass is 42,296.1 Da) (Figure 5.4-B). Three additional small peaks were also observed due to MALDI matrix ions causing interference.

¹Mutagenesis was performed by Carla Brackstone (Department of Chemistry, University of Warwick), while I designed the experimental side (design of primers, etc) and analysed the data.

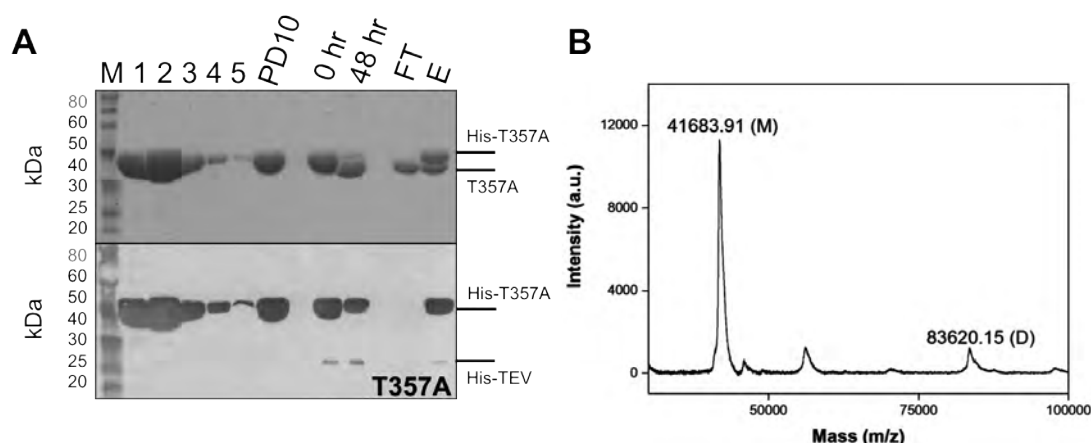


Figure 5.4: Typical purification of a mutant construct. *A)* Coomassie-stained SDS-PAGE gel (top) and anti-His western blot (bottom) summarising purification and cleavage of the His-tag from T357A. Lanes 1-5 show elution of the protein from a Ni-NTA affinity column, followed by desalting of the protein using a PD10 column, and cleavage of the His-tag from CPG2 using TEV protease. Cleavage was approximately 50% complete after 48 hours, at which point a second round of nickel affinity chromatography was used to collect the untagged T357A in the flow through (*FT*) and first elution fraction (*E*). A protein marker (*M*) is shown for reference. *B)* The MALDI-TOF mass spectrum of unlabelled T357A. Monomeric and dimeric species were observed in the mass spectrum (calculated size: 42,296.1 Da).

The same procedure was successfully applied for purification of H112A, H385A, E175A, E200A, D141A, K208A and S210A mutants (Figure 5.5). For MALDI-TOF analyses, a positive ion linear mode method was used to collect the mass spectra. In this mode, MALDI produces positive ions, usually in the form of $(M+H)^+$. The technique can also produce multiply charged ions, usually up to +3, as well as dimers, trimers, *etc.* Figure 5.5 shows that dimeric species are observed in the mass spectra of mutants H385A, E175A, E200A, K208A and S210A, although none of the SDS-PAGE gels, nor western blots, suggested the presence of higher order oligomeric state for any of these mutants.

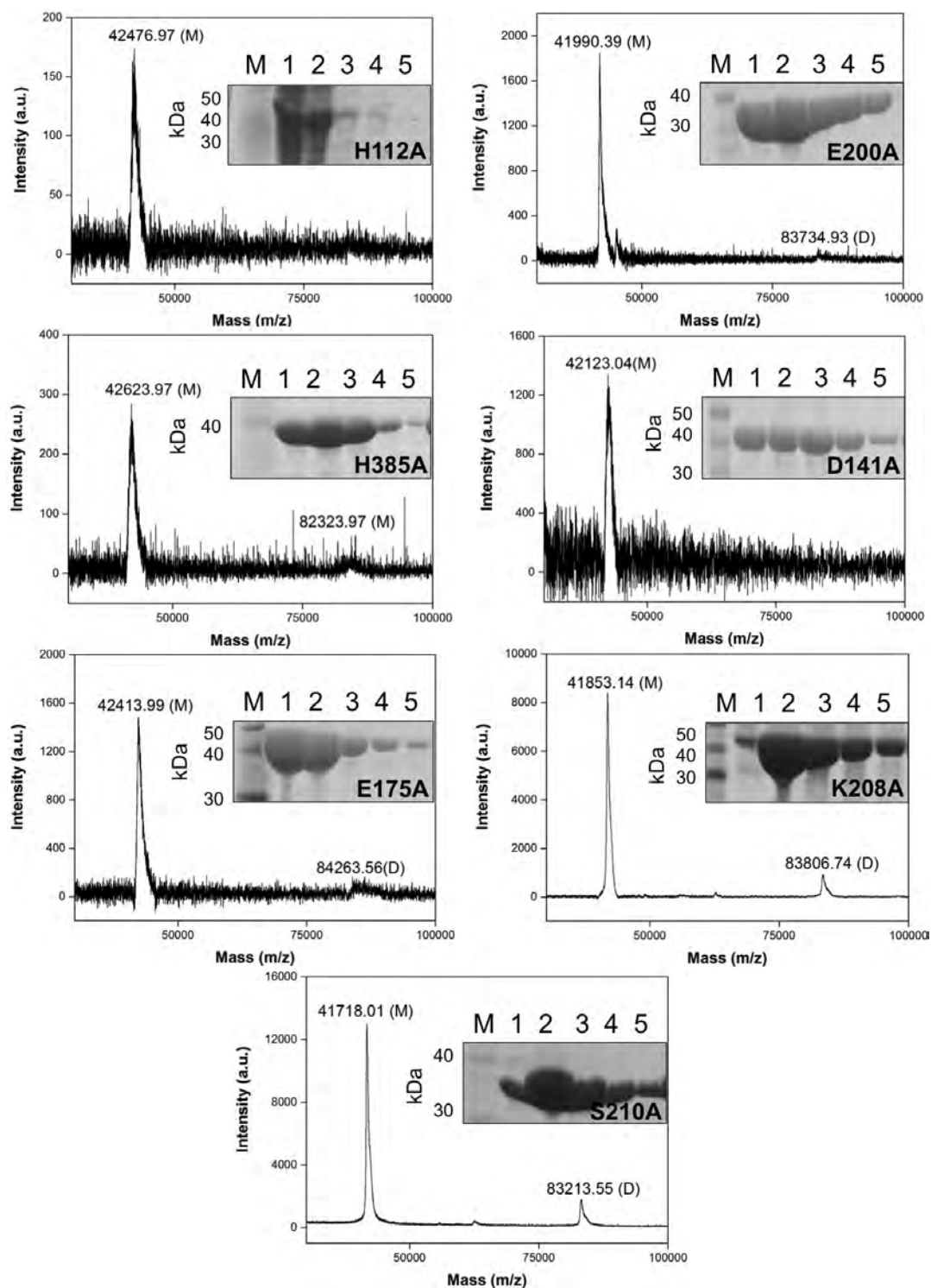


Figure 5.5: Purification of mutants. The same purification protocol used for T357A mutant shown in figure 5.4 was applied for the rest of the CPG2 mutants. The coomassie-stained SDS-PAGE gel purification is shown as an inset to the MALDI-TOF mass spectrum. Lanes 1-5 show elution of the protein from a Ni-NTA affinity column. A protein marker (M) is shown for reference. Both monomeric and dimeric species were observed in the mass spectrum of mutants H385A (calculated size without His-tag: 42,260 Da), E175A (calculated size without His-tag: 42,260 Da), E200A (calculated size without His-tag: 42,268.1 Da), K208A (calculated size without His-tag: 42,269 Da) and S210A (calculated size without His-tag: 42,310.1 Da).

To check whether these various point mutations had an impact on the expression levels of the mutant proteins compared to wild-type CPG2, relative quantification of gene expression was achieved by comparison of the expression level of the protein of interest with that of wild-type CPG2. Cell density for each sample was normalised to an OD_{600} of 6 to allow comparison. Figure 5.6 shows that the H112A mutant had the lowest expression level, followed by K208A and S210A mutants. On the other hand, H385A displayed the highest level of expression amongst all mutants. The protein yields (per litre of culture) for each mutant were calculated via BCA assay. The yield of all purified mutants protein was reduced in comparison to wild type expression levels (251.6 mg, Table 3.1 on page 82). H112A, D141A and K208A mutants had the lowest yields with 23.4 mg, 15.1 mg and 63 mg respectively, while H385A and E200A mutants displayed the highest expression yields with 153.1 mg and 125.2 mg respectively. In general, lower yield of purified mutant proteins could be attributed to the instability of the mutant proteins.

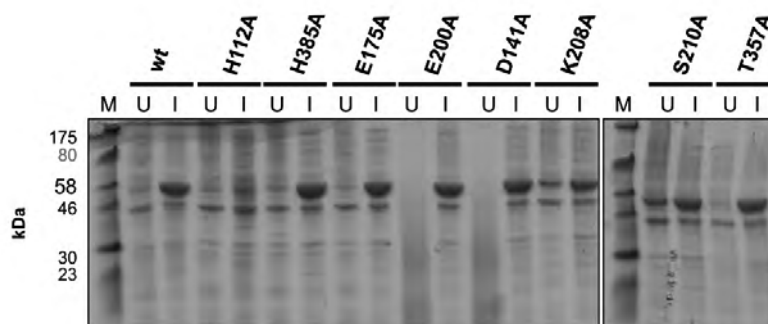


Figure 5.6: Coomassie-stained SDS-PAGE gel comparison of expression levels of wild type and its point mutant proteins. All point mutant plasmids were transferred in *E. coli* BL21 DE3 bacterial strain, where the recombinant mutant proteins were tested for expression before (*U*) and after (*I*) induction with 0.5 mM IPTG. The top band in each lane corresponds to the band of interest, and is absent from uninduced (*U*) lanes, except for K208A and S210A mutants where leaky expression was noticeable. Collected samples were normalised to an OD_{600} of 6 for comparison. The protein yields (per litre of culture) for H112A, H385A, E175A, E200A, D141A, K208A, S210A and T357A were 23.4 mg, 153.1 mg, 74.9 mg, 125.8 mg, 15.1 mg, 63.1 mg, 114.8 mg and 112.1 mg, respectively. The final concentration was determined using the BCA assay.

5.3 Investigation of Mutant Proteins Secondary Structure Using Circular Dichroism.

5.3.1 Secondary Structure Differences.

The CD spectra of the mutants produced were measured in the far-UV region in order to determine whether any important changes in secondary structure were observable in comparison to the native wild-type protein (Figure 5.7-A) (Greenfield 2006). CD measurements were performed at pH 7.3 in 20 mM Tris buffer. Far-UV CD spectra of the proteins at 25°C have a similar shape, with a negative maximum at 209 nm and positive maximum at approximately 190 nm, and similar intensities. The CD data at 25°C were fitted using the CDSSTR algorithm at DichroWeb and the percentage secondary structure content were estimated and are shown in Figure 5.7-B (Whitmore and Wallace 2008). The data showed very similar population-weighted combinations of each contributing secondary structure types, α -helix, β -sheet, turn and random coil, since these single amino acid mutants have overlapping spectra. This finding indicated that the mutant proteins have very similar secondary structure content to wild-type protein.

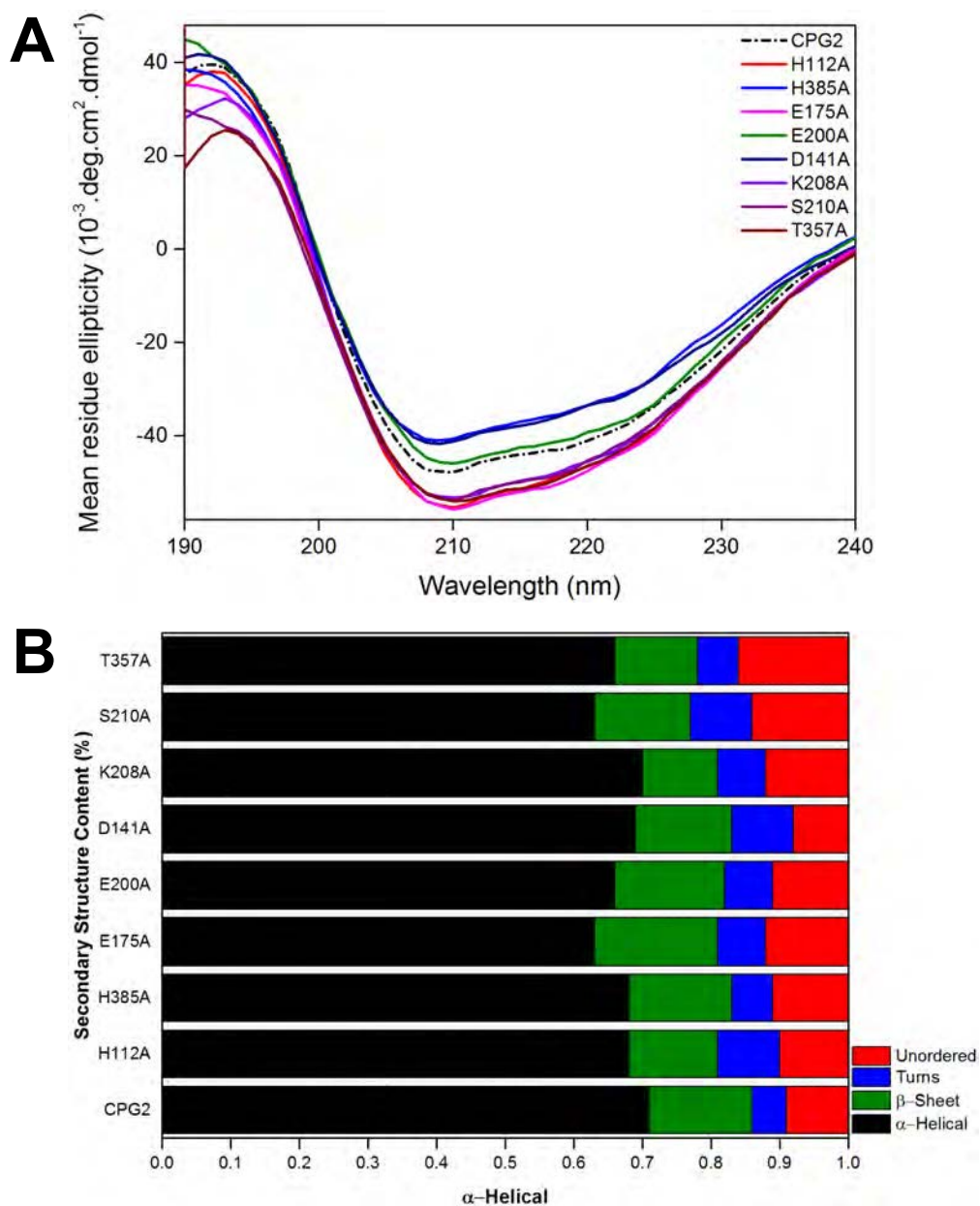


Figure 5.7: Circular dichroism. A) Far-UV CD spectra collected between 190-240 nm for wild-type CPG2 and all mutants (0.1 mg.mL^{-1}) solubilised in 20 mM Tris buffer containing 100 mM NaCl and 0.2 mM ZnCl_2 . B) Percentage of secondary structure for wild-type CPG2 and all mutants using DichroWeb (Whitmore and Wallace 2004). Circular dichroism measurements were recorded using a Jasco J-1500 spectrophotometer with 0.1 mm cylindrical cell. Each spectrum was accumulated at least 8 times at room temperature.

5.3.2 Thermal Stability.

Thermal stability of the wild type and mutant proteins was studied by monitoring the loss in secondary structure as a function of temperature by CD measurements. A wavelength of 220 nm was chosen to monitor α -helical features of the proteins structure. The CD signal at that wavelength was then recorded continuously as the temperature was raised from 20°C up to 90°C using a Peltier unit.

The results are collected in Figure 5.8 as scattered plots fitted (in red) to the sigmoidal boltzmann Equation (5.1) when possible.

$$y = A_2 + \frac{(A_1 - A_2)}{1 + e^{\frac{x-x_0}{dx}}} \quad (5.1)$$

where A_1 is the initial ellipticity value at 220 nm, A_2 is the final ellipticity value at 220 nm, x_0 is the T_m at which temperature is halfway between the initial and the final ellipticity value at 220 nm, dx describes the steepness of the curve.

All proteins aggregated or precipitated quickly after they were melted at 90°C, making unfolding irreversible. Melting temperatures for H385A, E200A, D141A, K208A, S210A and T357A mutants ranged from 53.24 to 64.28°C, which are very close to the wild-type protein T_m of 60.43°C. H112A showed the lowest T_m value of 28.23°C, while E175A showed the highest T_m value of 71.45°C. This implied that the E175A mutation conferred increased stability to the protein. On the other hand, H112A mutation decreased protein stability as it unfolded at a lower temperature.

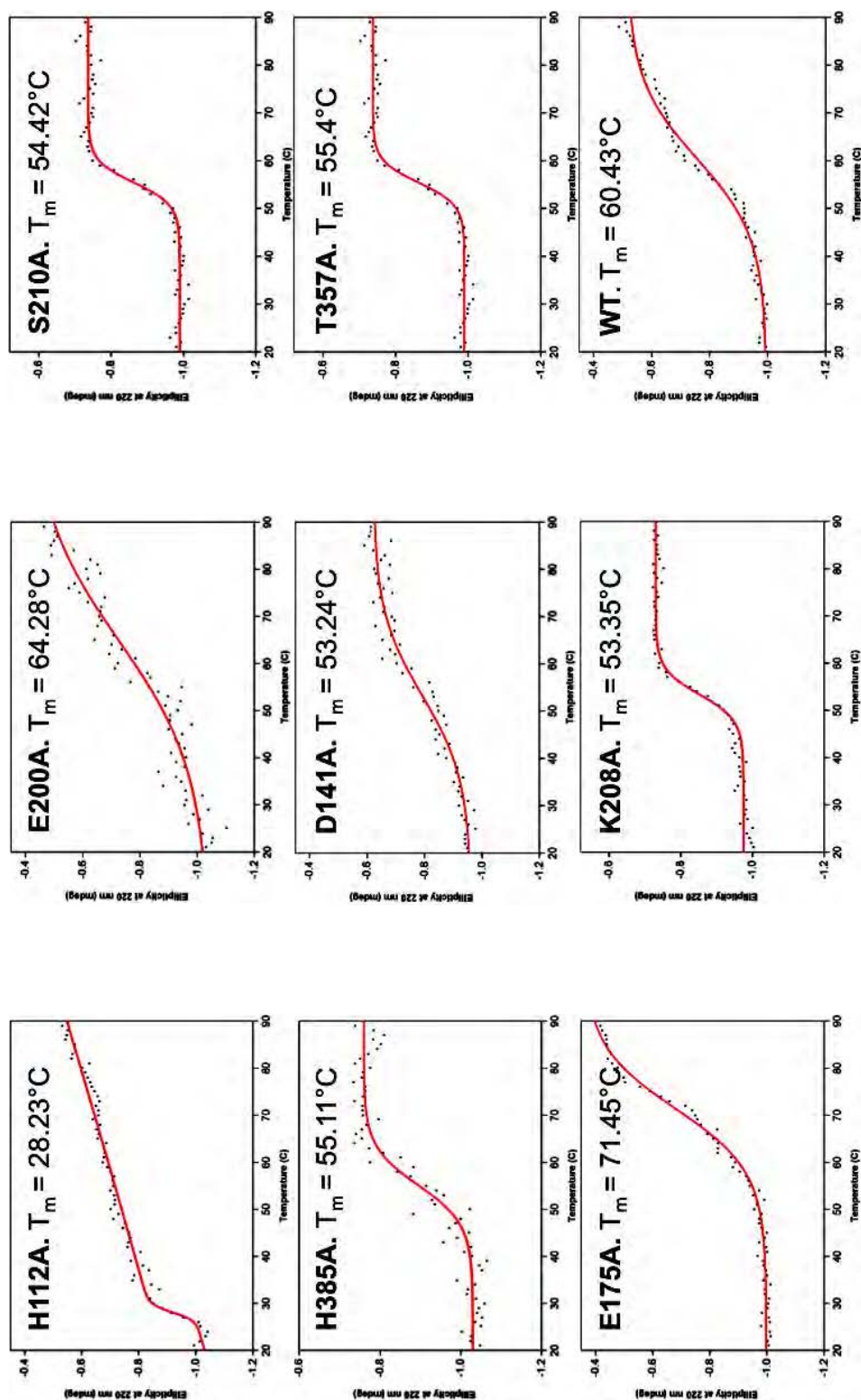


Figure 5.8: Melting temperature determined by CD melts at 220 nm. Circular dichroism measurements were recorded using a Jasco J-1500 spectrophotometer with 0.1 mm cylindrical cell using a Peltier temperature control unit. The CD data points were fitted with a Boltzmann sigmoidal function for all mutants (except for H112A, which was fitted with a double Boltzmann sigmoidal function) and wild type CPG2 using Origin 9.1. Calculated T_m values for each mutant (and wild-type) are displayed within each corresponding graph.

The unfolding mechanism for many globular proteins is essentially a two-state process, where only the folded and unfolded states are reasonably populated at equilibrium (Pace et al. 1999; Waldner et al. 1999). This was observed for wild-type, H385A, E175A, E200A, D141A, K208A, S210A and T357A mutant proteins where proteins from a folded state, at lower temperatures, continued onto an unfolded state as temperature was increased. The presence of a sigmoidal curve is usually considered as a strong indication of the occurrence of a two-state denaturation process (Privalov and Khechinashvili 1974; Privalov 1997). The cooperativity of the unfolding reaction is measured qualitatively by the width and shape of the unfolding transition (Consalvi et al. 2000). A rapid and highly cooperative unfolding reaction was observed for H385A, S210A, K208A and T357A (Figure 5.8-H385A, S210A, K208A and T357A). In the X-ray crystal structure, these residues are located in the dimerisation and catalytic domains of CPG2. The data suggest that these proteins exist initially as a compact well-folded structure. The mutation of these residues from the dimerisation and catalytic domains seemed to also increase the rate of unfolding, which suggested that both domains may unfold as a single, larger cooperative unit compared to wild-type (Vermeer and Norde 2000). It is inferred that the H385A, S210A, K208A and T357 mutants may increase the stability of the domain they are respectively located in, and / or the neighbouring domain. Secondly, a very gradual, non-cooperative melting reaction was observed for E175A, E200A, D141A and wild-type proteins (Figure 5.8-E175A, E200A, D141A and WT), all of which are located in the catalytic domain. This suggested that the proteins existed initially as a very flexible,

folded protein, which is consistent with the presence of the two different domains (i.e catalytic and dimerisation domains) in the protein. The mutation of these residues in the very flexible region of the catalytic site may have not significantly changed the cooperative unfolding behaviour of the two domains within the protein, compared to wild-type. However, H112A (located in the catalytic domain) displayed a peculiar behaviour compared to the other mutant and wild-type proteins (Figure 5.8-H112A). The H112A melting curve showed a double sigmoidal curve, as if the H112A mutant protein could not unfold in a single transition, but in two transitions. This may indicate that the catalytic and dimerisation domains of H112A unfolded independently, with one of the domains unfolding rapidly at low temperatures (around 30°C) compared to the other domain, where a slow unfolding transition was observed at higher temperature (Vermeer and Norde 2000). It may also suggest that mutation of residue H112 in the catalytic site had a bigger impact on the cooperative unfolding behaviour of the two CPG2 domains than the mutation of residues E175, E200 and D141, also located within the catalytic site, had. As a result H112A melting curve could only be fitted using the double Boltzmann equation in order to obtain the T_m of the first unfolding domain.

5.4 Use of CPG2 Spectrophotometric Assay to Study the Effects of Residue Mutation on Substrate Binding.

Finally a spectrophotometric assay was used to measure the catalytic activity of the mutant proteins, by monitoring the rate of degradation of MTX at 320 nm. Figure 5.9 shows the results. We expected that the mutations disturbing the catalytic site would inevitably result in inactive forms of CPG2, whereas the mutations involving P1 pocket would result in reduced activity of CPG2. H112A, E175A and D141A mutant proteins (found in the catalytic domain) revealed very low activities against MTX. However, H385A and E200A showed measurable activity. Although the activity is reduced compared to wild-type CPG2, it was noteworthy. H385A and E200A showed activities of 50.87 and 63.70 U/mg, respectively. Residue H385 co-ordinates zinc 1, whereas amino acid E200 co-ordinates zinc 2 (Figure 5.1). H112A mutant protein resulted in inactivity, while its counterpart H385A co-ordinating zinc 1 appeared active (Figure 5.3). H112A mutant protein was less active at low temperature than H385A mutant protein, which could explain activity differences. The same comparison between E200A and E176 could not be made as the pet151-E176A plasmid would not transform into BL21 DE3 *E. coli* strain. There were possibly other residues in the region that have compensated for the absence of residues H385 and E200. Mutants K208A, S210A and T357A also had very low-activity against MTX. It implied that these residues, residing in P1 pocket, are possibly

involved in substrate recognition.

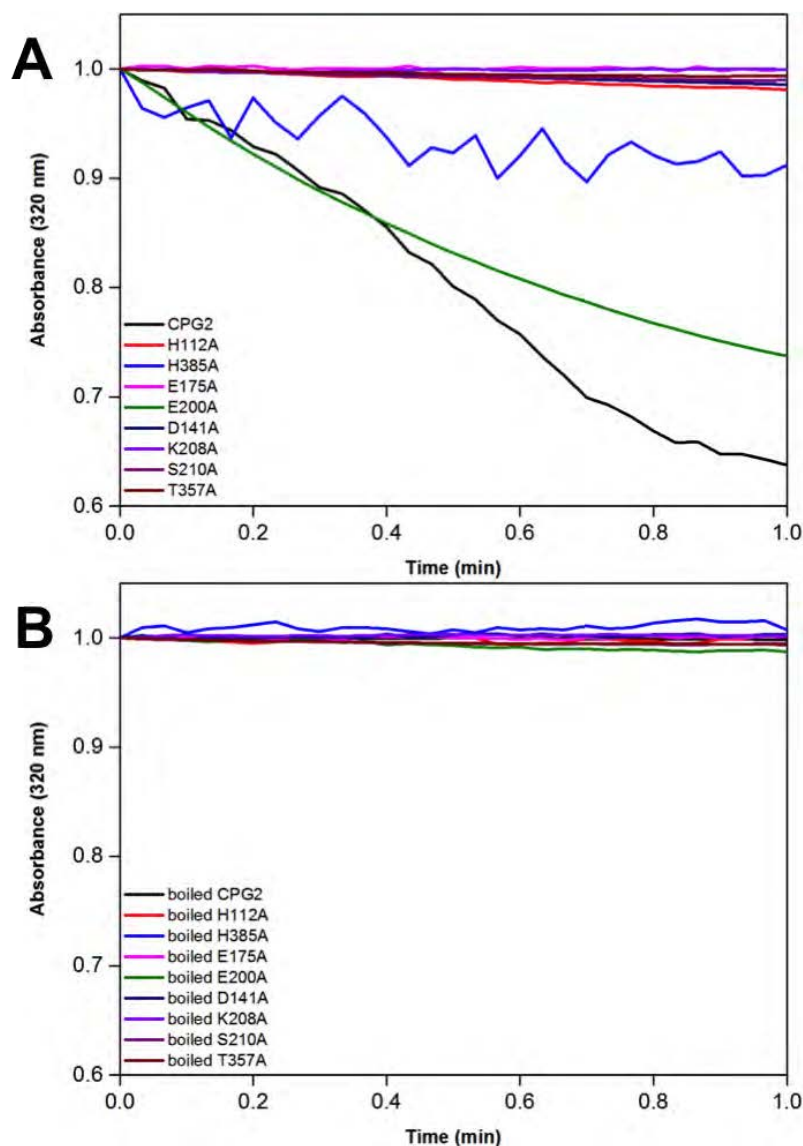


Figure 5.9: Enzymatic activity assays of each mutant and wild-type proteins. *A)* Enzymatic activity measurements for all mutants as measured using the methotrexate cleavage assay. *B)* Enzymatic activity measurements for boiled protein was included as a negative control in the same assay. The calculated specific activity values from these assays were CPG2: 96.7 U/mg, H112A: 3.91 U/mg, H385A: 50.87 U/mg, E175A: 2.40 U/mg, E200A: 63.70 U/mg, D141A: 1.75 U/mg, K208A: 3.93 U/mg, S210A: 1.46 U/mg and T357A: 1.42 U/mg.

The next step was to analyse the sensitivity of mutants that remained active after mutation to CP06 and CP65 inhibitors (Figure 4.1 on page 112). To test a potential inhibitor of CPG2, it was diluted with the assay buffer. The final concentration of the solution depended on the solubility of the material in the buffer. Both H385A and E200A mutant proteins were active against MTX, and

inhibited by CP06 and CP67 (Figure 5.10). This suggested that CP06 and CP67 bind to the CPG2 active site, preventing MTX from binding and being degraded by CPG2. This data also indicated that H385 and E200 residues can be replaced by an alanine residue and still behave very similarly to wild type CPG2 protein.

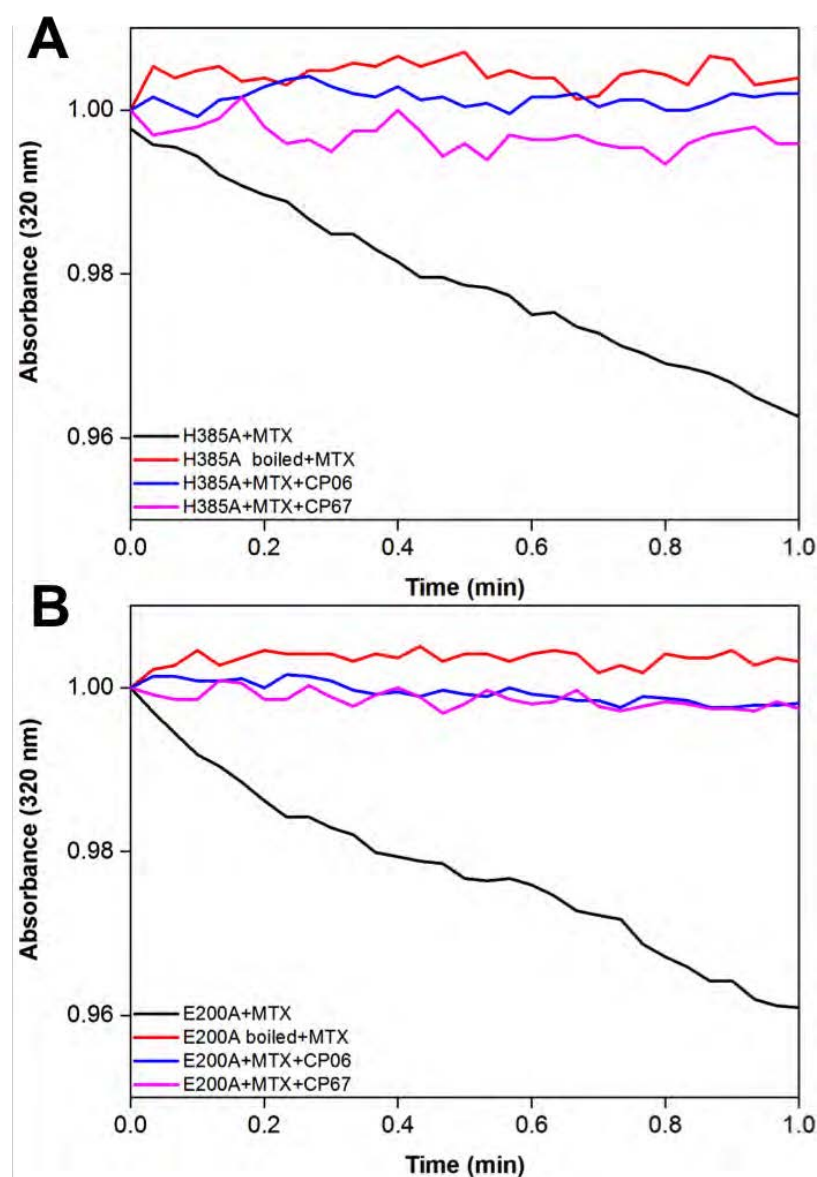


Figure 5.10: Competitive enzymatic activity assays of H385A and E200A mutant proteins. *A)* Competitive enzymatic activity measurements for H385A mutant using CP06 and CP67 inhibitors as measured using the methotrexate cleavage assay. *B)* Competitive enzymatic activity measurements for E200A mutant using CP06 and CP67 inhibitors as measured using the methotrexate cleavage assay.

5.5 Conclusion.

The main question concerned whether the selected residues for site-directed mutation revealed the potential binding site of MTX and their inhibitors on CPG2. First, all constructs were easily made, transformed in *E. coli* BL21 DE3, expressed and purified for characterisation, except for E176A and G360A (Figures 5.4 and 5.5). As a result, P2 pocket's implication in substrate recognition could not be clarified.

Characterisation of the proteins expressed revealed the presence of dimers in the MALDI-TOF spectra for H385A, E175A, E200, K208 and S210A mutant proteins. This was also observable with the wild-type protein (Figure 3.10-B, page 92). This suggested that the mutation of these residues did not affect the capacity for the protein to dimerise.

Moreover, the effect of a mutation on the secondary structure of a protein can be studied by CD spectroscopy in the "far-UV" spectral region (190-240 nm) (Whitmore and Wallace 2004). Unfortunately, CD analysis did not present considerable discrepancies in the far-UV region (Figure 5.7). Therefore, it was decided to study stability of the protein using CD thermal melts (Figure 5.8). For most proteins secondary structure is lost upon unfolding and the far-UV CD spectra of a folded and unfolded protein are therefore unique. Thus, the thermal stability of each mutant and wild-type proteins was investigated by measuring the intensity of the CD signal at a fixed wavelength (e.g. 220 nm for the α -helical region) as a function of temperature. Melting temperatures for H385A, E200A, D141A, K208A, S210A and T357A mutants ranged from 53.24

to 64.28, which are very close to the wild-type protein T_m of 60.43°C. Additionally, the shape of the sigmoidal curve allowed to understand more about the unfolding process of each protein. H385A, E175A, E200A, D141A, K208A, S210A, T357A and wild-type proteins were in a folded state at lower temperatures which continued onto an unfolded state as temperature was increased. This implied that the unfolding process for these proteins was essentially a two-state denaturation process. The cooperativity of the unfolding reaction was measured qualitatively by the width and shape of the fitted sigmoidal curve (Figure 5.8). A rapid and highly cooperative unfolding reaction was observed for H385A, S210A, K208A and T357A. It suggested that the protein existed initially as a compact, well-folded structure. Secondly, a very gradual, non-cooperative melting reaction, was observed for E175A, E200A, D141A and wild-type proteins indicating that the proteins existed initially as a very flexible, folded protein, which is consistent with the presence of the two different domains (i.e catalytic and dimerisation domains) in the protein. Amongst, all these proteins only H385A, E200 and wild-type demonstrated activity against MTX substrate (Figure 5.9). H385 showed the lowest activity of 50.87 U/mg, whereas both wild-type and E200A presented activities of 96.70 and 63.70 U/mg, respectively. The greater specific activity levels may be also associated with higher flexibility observed within the folded wild-type and E200A proteins. However, the melting data showed significant difference in T_m with H112A and E175A construct. H112A showed the lowest T_m value of 28.23°C, while E175A showed the highest T_m value of 71.45°C. H112A mutant protein also showed a different tendency. The H112A melting curve showed a

double sigmoidal curve, as if H112A protein could not unfold in a single transition, but in two transitions. This is also consistent with the presence of two different domains within the protein and suggested the unfolding of one of the domains in preference to the other. Since H112 residue is found in the catalytic domain, it may imply that the catalytic domain was the first to unfold and its secondary structure became more unstable upon mutation. These data were supported by the low expression level observed showed in Figure 5.6 and the absence of activity against MTX observed in Figure 5.9.

To conclude, only two mutations within the catalytic site, H385A and E200A showed activity against MTX and were inhibited by CP06 and CP67. This suggested that not only CP06 and CP67 were competitive CPG2 inhibitors, but that H385 and E200 residues did not play an important part in the MTX recognition process because replacement of both residues by alanine still generated active proteins. However, inhibition of H385A and E200A mutant protein was achieved by CP06 and CP67. Since these residues are co-ordinated to both zinc ions, these data indicated that the protein may have compensated for their absence some other way. H385A showed better stability than wild-type, whereas E200A showed higher flexibility than the wild-type CPG2 according to CD thermal melts analysis. All other mutants showed loss of activity. E175A, E200A and D141A mutants had reduced activity against methotrexate implying that they were crucial for catalysis to occur, as they co-ordinate to the metal centres. However, there is no clear explanation as to why H112A resulted in inactivity, when its counterpart H385A co-ordinating zinc 1 appeared active (Figure 5.3). Although residues in P2 pocket could not

be analysed, K208A, S210A and T357A mutations from P1 pocket manifested no activity against MTX, validating their involvement in substrate recognition.

Chapter 6

Characterisation of Engineered Individual Domains of CPG2

6.1 Introduction.

We have seen in previous chapters that one of the most important tools for probing physiological processes, mainly controlled by intermolecular recognition mechanisms involving protein-ligand interactions, is solution state NMR spectroscopy through protein-based and ligand-based experiments.

In protein-based methods, chemical shift is the easiest variable to monitor in order to locate the ligand binding site. The major disadvantages of protein-based methods are the experimental time, and the need for isotopically-labelled protein with good solution properties soluble at high concentration. Chapter 3 describes methods used to overcome these challenges. Unfortunately, these methods are

restricted to proteins with relatively low molecular masses (below 30 kDa) to avoid difficult labelling strategies and full resonance assignment. In order to explore the structure of CPG2 in solution upon binding of its inhibitors, the NMR resonances must be assigned to specific nuclei of the protein. As the size of the protein increases, the assignment process becomes progressively problematic because of amplified spectral overlap in the already poorly resolved NMR spectra. Efforts at more exhaustive labelling ($^{15}\text{N}/^{13}\text{C}/^2\text{H}$) were unable to overcome this issue; hence a “divide-and-conquer” approach where individual domains of the full-length enzyme were cloned, expressed and purified was used to reduce the molecular-weight to a range more amenable for NMR studies.

6.2 Different Domains of CPG2.

According to the crystal structure (PDB ID: 1CG2), CPG2 consists of distinct protein domains associated with different functions, namely a dimerisation domain inserted between two non-contiguous regions of the catalytic domain (Rowse et al. 1997). Rowse et al. (1997) suggested that the catalytic domain is likely to contain the binding site as it provides the substrate with two zinc ions in the active site required for cleavage. Considering this molecular domain organisation, we decided to dissect the protein into three smaller fragments suitable for NMR structural studies: the 12.73 kDa dimerisation domain fragment (CPG2_{DIM}, residues 213-322); the full 30.95 kDa catalytic domain fragment (CPG2_{CAT}, composed of residues 23-214 and residues 323-415 linked via a non-native alanine); and the largest 21.27 kDa contiguous region of the

catalytic domain (CPG2_{CAT}, composed of residues 23-198) (Figure 6.1).

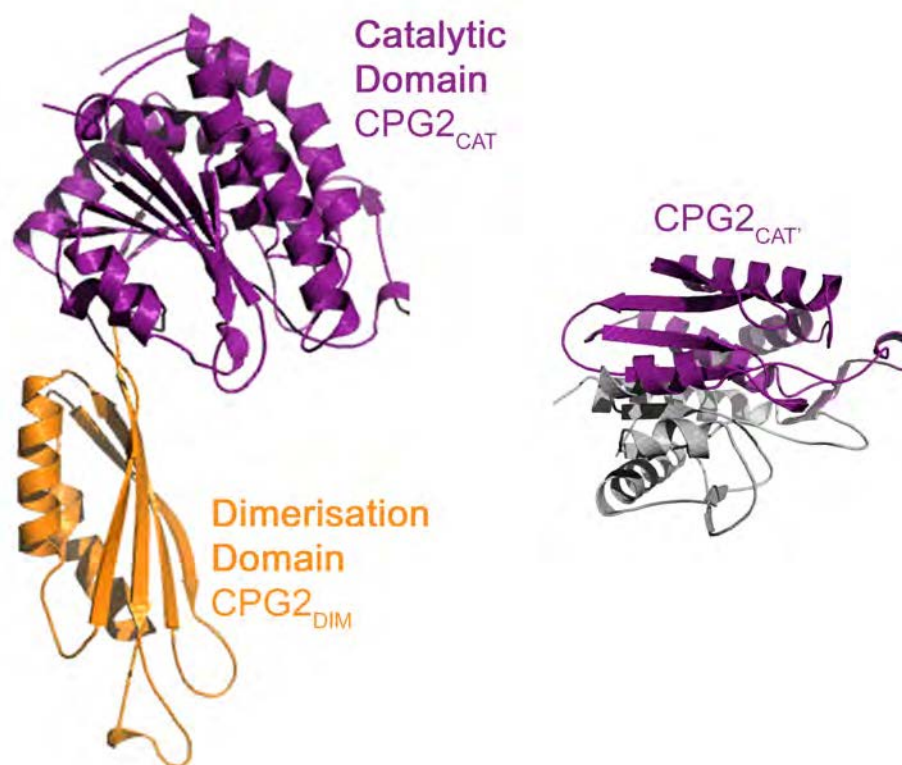


Figure 6.1: Different domains of CPG2. Left panel: coloured in purple is the catalytic domain and in orange is the dimerisation domain of CPG2. Right panel, the largest contiguous region of the catalytic domain coloured in purple is shown. PDB: 1CG2, (Rowse et al. 1997).

6.3 Cloning and Expression of the Individual Domains.

6.3.1 Preparation of Plasmid DNA Constructs.

The pET151-D-topo plasmid containing the CPG2₂₃₋₄₁₅ gene was used as a template to generate cDNA sequences of the different constructs (CPG2_{DIM},

CPG2_{CAT} and CPG2_{CAT'}), which were then sub-cloned into pET151-D-TOPO vector (Invitrogen, UK) with an N-terminal tag containing the V5 epitope and a hexahistidine affinity tag, which could be removed by TEV cleavage. The CPG2_{DIM} and CPG2_{CAT'} constructs were easily generated, while the CPG2_{CAT} (made of two non-contiguous regions of the catalytic domains) required two regions of the gene to be amplified and ligated through the presence of an extra non-native amino acid, alanine (Figure 6.2).



Figure 6.2: Construction of His-tagged recombinant proteins. Schematic diagram shows primary structure (N- and C-termini of protein) and functional domains (signal peptide: amino-acids 1-23, catalytic domain: amino-acids 23-214 and 323-415, and dimerisation domain: amino-acids 213-322) of CPG2 as described previously.

It was unknown whether the absence of the dimerisation / catalytic domains, or the addition of an extra alanine residue between the two non-contiguous regions of the catalytic domain, would have an impact on the fold of the different domains generated (Figure 6.1). All the constructs were successfully prepared, confirmed via multiple analyses of PCR products performed on 1-2% agarose gels and DNA sequencing by (GATC, Germany) as described in Figure 6.3.

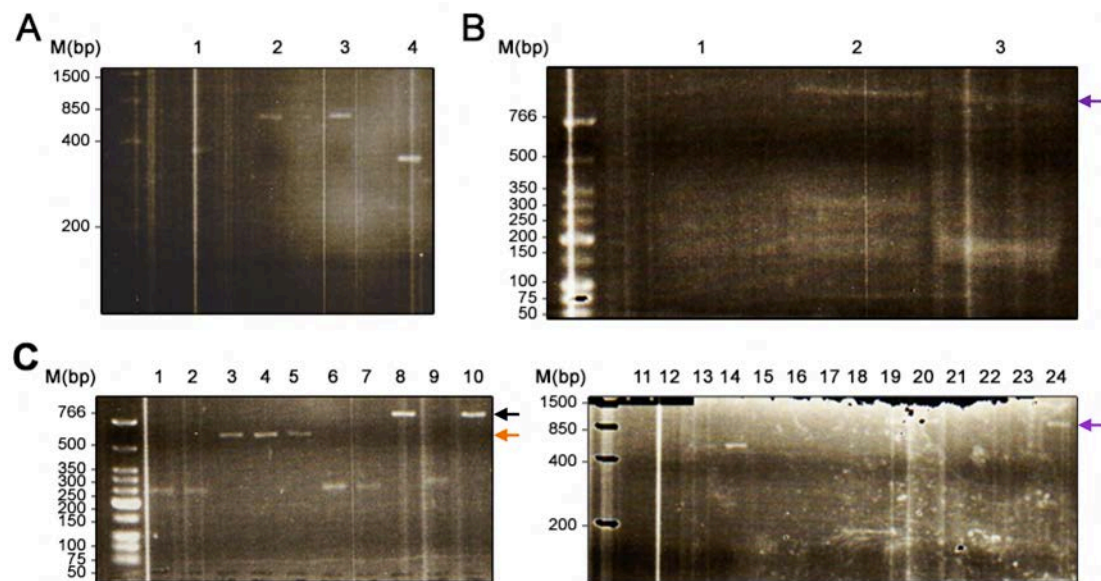


Figure 6.3: 1-2% agarose gel electrophoresis of PCR products. A) PCR products from pet151-CPG2₂₃₋₄₁₅ plasmid using the appropriate pair of primers for each construct as described in Table 2.3 of Chapter 2. Lane 1: amplification of the CPG2_{DIM} gene (expected band: 330 bp). Lane 2: amplification of the gene encoding the N-term region of the CPG2_{CAT} domain (expected band: 528 bp). Lane 3: amplification of the largest sequence of CPG2_{CAT} gene (expected band: 576 bp). All PCR products were confirmed by sequencing. Lane 4: amplification of the gene encoding the N-term region of the CPG2_{CAT} domain (expected band: 282 bp). B) Lanes 2-3: Positive PCR products (indicated with a purple arrow) resulting from the ligation of the largest and smallest genes composing CPG2_{CAT} gene sequence. C) Cloning confirmation of the gene of interest into pET151 vector by colony PCR using T7 forward and reverse primers. Lanes 3-4: positive PCR products for pet151-CPG2_{DIM} (expected band: 592 bp, indicated with an orange arrow, left panel). Lanes 8 and 10: positive PCR products for pet151-CPG2_{CAT} (expected band: 790 bp, indicated with a black arrow, left panel). Lanes 19, 20 and 24: positive PCR product for pet151-CPG2_{CAT} (expected band: 1120 bp, indicated with a purple arrow, right panel).

6.3.2 Effect of Changes in Temperature and IPTG Concentration on Expression.

To investigate whether these domains were able to fold and function independently of each other, each of the above three constructs was expressed with uniform labelling in *E. coli*. Induction temperatures and IPTG concentrations were varied to optimise the expression levels in rich media. Figure 6.4 shows that optimal expression was achieved at 25°C for all three proteins. Changes in IPTG concentration had little impact on expression levels of CPG2_{CAT} and CPG2_{DIM} at 25°C (Figures 6.4-A,C). However, upon visual inspection of the SDS-PAGE gel, expression levels of CPG2_{CAT} improved when 0.7 mM IPTG was added (Figure 6.4-B). Hence, to keep the expression protocol identical across all constructs, for simplicity, protein expression was induced with 0.7 mM IPTG induction at 25°C for 18 hours.

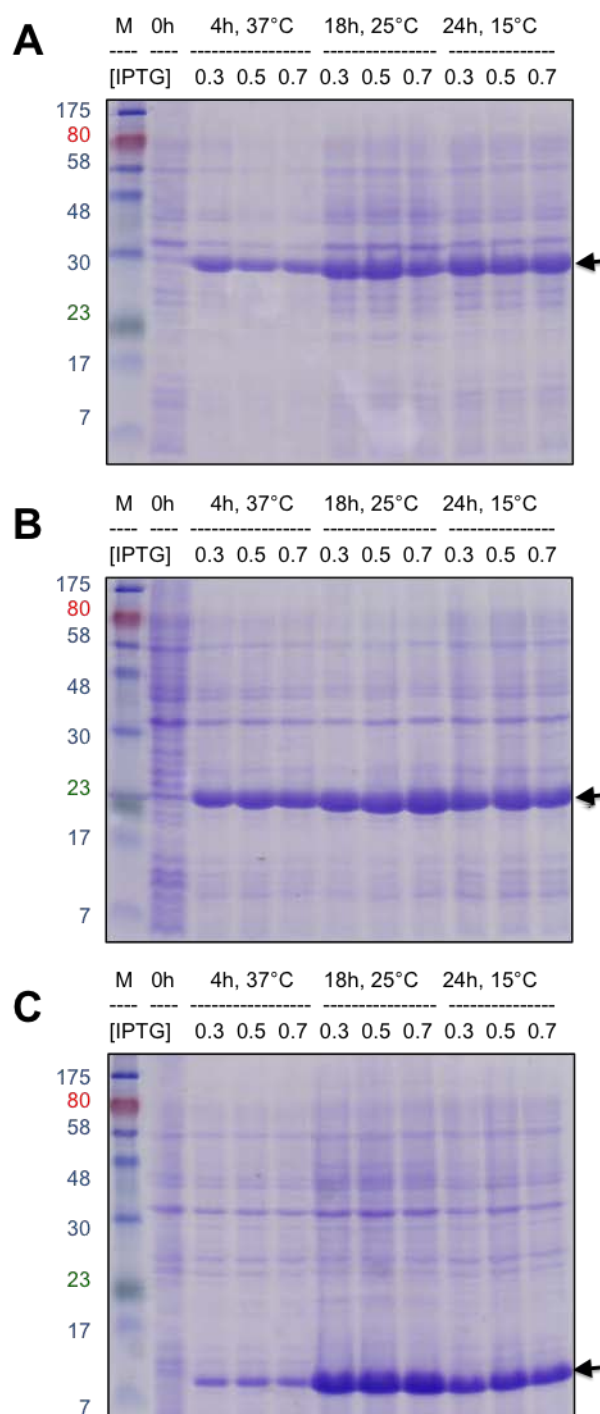


Figure 6.4: Effects of IPTG concentration and temperature changes on expression. Total expression levels of *A*) CPG2_{CAT}, *B*) CPG2_{CAT'} and *C*) CPG2_{DIM} at 37°C, 25°C and 15°C with 0.3 mM, 0.7 mM and 0.9 mM IPTG (expression band of the corresponding protein of interest is indicated with a black arrow).

6.4 Preparation of Isolated CPG2 Domains

Narrows Down Requirements for Activity.

To streamline expression of multiple proteins in parallel, the optimal expression conditions in rich media were directly applied to expression of CPG2 fragments in labelled minimal media. The condensation method (see in Section 2.2.5 of Chapter 2) was also applied, but without the use of double colony selection. As shown in Figure 6.5 all three constructs were purified using the same purification protocol used for CPG2₂₃₋₄₁₅ in order to avoid additional optimisation steps for purification (Section 3.2). The activity of each labelled construct was then assessed using the methotrexate-based activity assay and the activities are provided in Table 6.1.

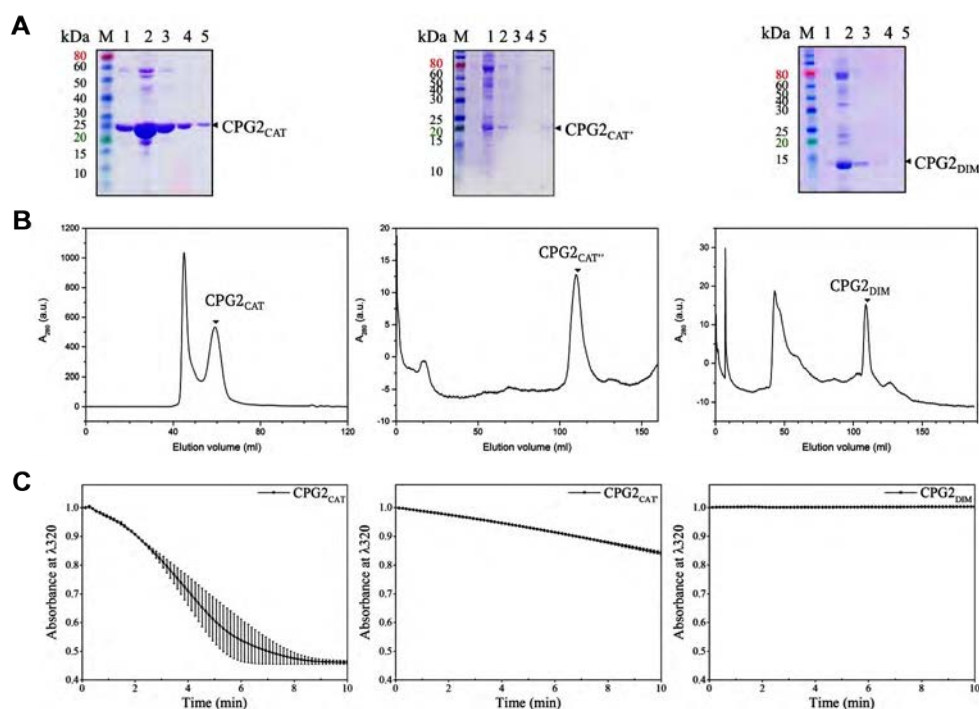


Figure 6.5: Characterisation of CPG2_{CAT}, CPG2_{CAT'} and CPG2_{DIM}. *A*): SDS-PAGE purification gels. (*M*) denotes the molecular weight marker, lanes (1) to (5) show protein samples eluted after purification by Ni-IMAC chromatography. *B*): Size-exclusion chromatography for (from left to right) [U-²H,¹³C,¹⁵N] CPG2_{CAT}, [U-¹⁵N] CPG2_{CAT'} and [U-¹⁵N] CPG2_{DIM} constructs respectively (the black arrow indicates elution of monomeric untagged protein). *C*): Enzymatic activity assay against the substrate methotrexate.

As expected, CPG2_{DIM} yielded low activity (compared to CPG2₂₃₋₄₁₅) (Table 6.1). CPG2_{CAT} yielded activity was 62% of that observed for CPG2₂₃₋₄₁₅, suggesting that linking these two polypeptide chains and adding an extra amino acid (alanine) did not disrupt the domain structure so critically that all function was lost (Table 6.1). Removal of residues 323-415 from the catalytic domain, to create the CPG2_{CAT'} protein, lead to a further reduction in catalytic activity indicating that the C-terminal region of the catalytic domain is implicated in substrate recognition / activity.

Table 6.1: Difference in expression yield between CPG2₂₃₋₄₁₅ and its individually expressed domains.

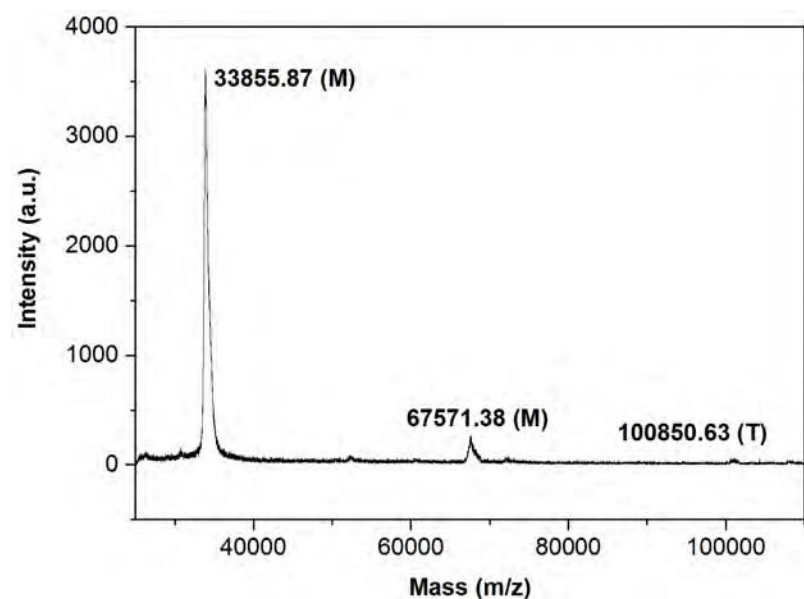
Constructs	His-tag Terminus	Activity U/mL	Specific Activity U/mg	Protein Yield [†]		
				Unlabeled (2×YT)	¹⁵ N (M9)	² H/ ¹³ C/ ¹⁵ N (M9)
CPG2 ₂₃₋₄₁₅	N	100.3	75.5	251.6	nd [‡]	70-109.42
CPG2 _{CAT}	N	62.8	42.7	91.4	65.8	95.2
CPG2 _{CAT'}	N	8.2	3.7	nd	21.6	nd
CPG2 _{DIM}	N	1.5	2.9	48.8	5	nd

[†] Yields of purified proteins are in mg/L of culture.

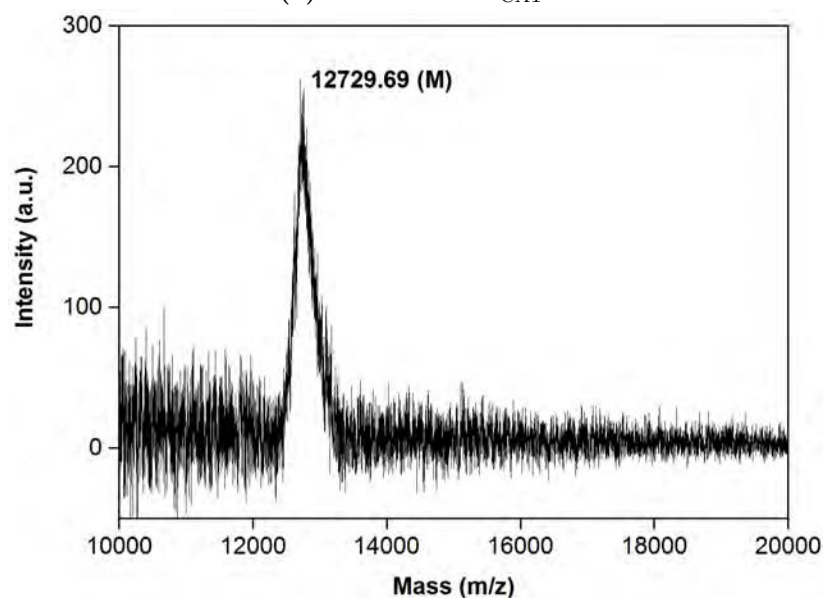
[‡] Not determined.

At the high concentrations required for NMR analyses, it was crucial to eliminate any ambiguity originated by the presence of dimers. Gel-filtration suggest the presence of higher oligomeric states for CPG2₂₃₋₄₁₅ (Figure 3.3-B) and CPG2_{CAT} (Figure 6.5-B), but no dimers (or higher order oligomers) were observed for CPG2_{DIM}, which was thought to form the dimer interface in

CPG2₂₃₋₄₁₅ (Figure 6.6-B). This is likely due to the low concentration of purified CPG2_{DIM} (Table 6.1). For the constructs that did form oligomers, gel filtration chromatography was used to isolate the monomeric form of these proteins (Figure 6.5-B).



(a) Purified CPG2_{CAT}.



(b) Purified CPG2_{DIM}.

Figure 6.6: MALDI-TOF spectra of the isotopically labelled proteins. A) [U-²H, ¹³C, ¹⁵N]-labelled CPG2_{CAT} (theoretical monomeric size: 35,275 Da). B) [U-¹⁵N]-labelled CPG2_{DIM} construct (theoretical monomeric size: 12,932 Da). M, D, and T denote the presence of monomer, dimer and trimer respectively.

6.4.1 Purification of CPG2_{CAT}, CPG2_{CAT'}, CPG2_{DIM} by Gel Filtration Chromatography.

Elution fractions from IMAC containing CPG2_{CAT}, CPG2_{CAT'}, CPG2_{DIM} were next applied to PD10 desalting columns (see Section 2.2.6) and further purified by Gel Filtration Chromatography (see Section 2.2.7). Improved separation of CPG2_{CAT}, CPG2_{CAT'}, CPG2_{DIM} proteins was achieved by applying samples that had been concentrated using 10 kDa MWCO (and 5 kDa MWCO for CPG2_{DIM}) concentrators (see Section 2.6.8) to a HiLoad 16/600 Superdex 75 g GL column (separation range 1-120 kDa), under an isocratic gradient of 100% A (20 mM Tris, 100 mM NaCl, pH 8). As shown in Figures 6.5-B and 6.7, CPG2_{CAT} eluted as a double peak at elution volumes of 65.99 mL for dimeric CPG2_{CAT} and 57.15 mL for monomeric CPG2_{CAT}. CPG2_{CAT'} and CPG2_{DIM} eluted as single peaks at elution volumes of 71.17 mL and 76.56 mL respectively (Figure 6.5-B).

By applying Gel Filtration standards to the column and running at the same flow rate and with the same buffer (see Section 2.2.7), it was possible to create a calibration curve of molecular weight against elution volume, assuming a void volume of 25.0 mL. Plotting the relative retention times of CPG2_{CAT} protein on this curve gave predicted molecular weights of around 34 kDa and 70 kDa (calculated monomeric peak for CPG2_{CAT} is 30,736 Da), and therefore CPG2_{CAT} is expected to migrate and exist as a monomer and a dimer, as a consequence it seem to retain CPG2₂₃₋₄₁₅ protein's capacity to dimerise. However, plotting the relative retention times of CPG2_{CAT'} and CPG2_{DIM} proteins on this curve gave predicted molecular weights of around 22 kDa and

14 kDa respectively (calculated monomeric peak for CPG2_{CAT'} is 19428.10 Da and calculated monomeric peak for CPG2_{DIM} is 12497.100Da), and therefore CPG2_{CAT'} and CPG2_{DIM} are expected to migrate and exist as monomeric species, therefore both fail to exist as dimers. This implies that the shortest region of the catalytic domain may be involved in the dimerisation process.

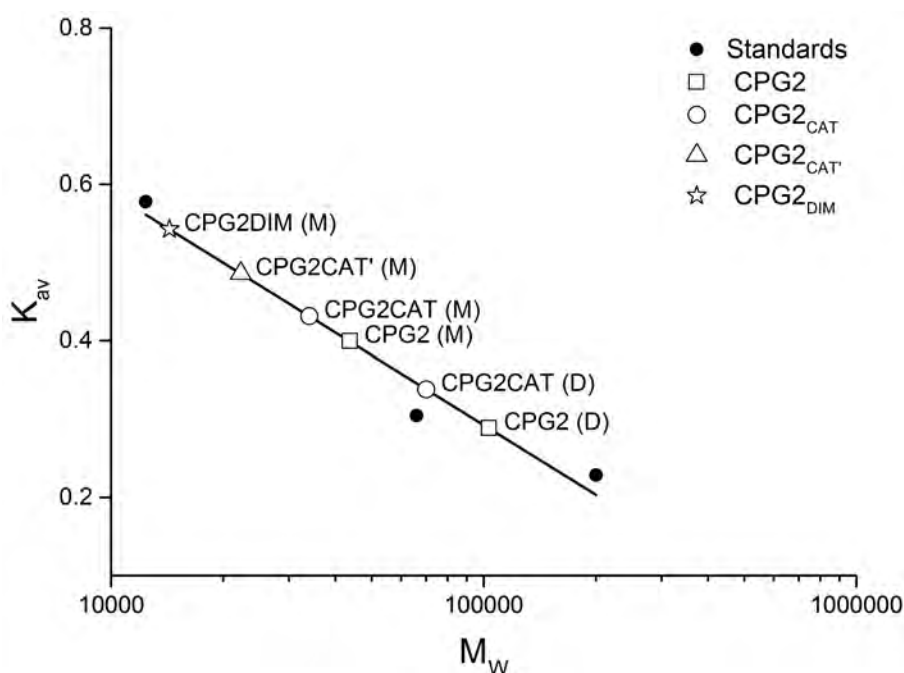
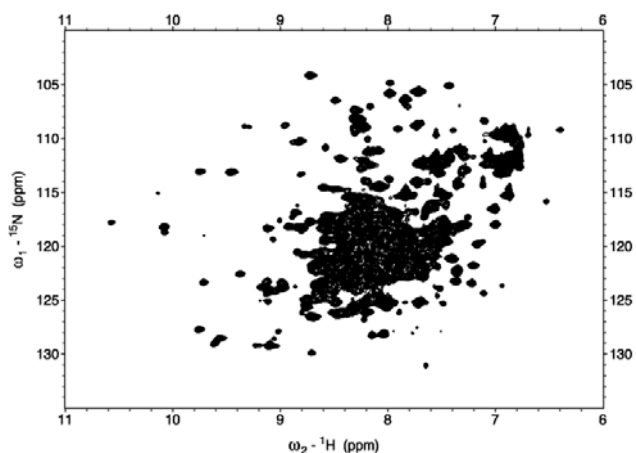


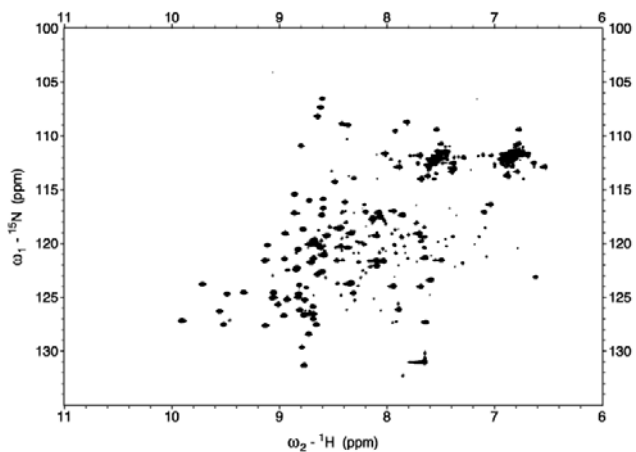
Figure 6.7: Calibration curve obtained with Gel Filtration standards on HiLoad 16/600 Superdex 75 g GL column. An isocratic gradient of 100% of buffer A (20 mM Tris, 100 mM NaCl, pH 8) was used. The calibration curve graph represents the gel phase distribution coefficient K_{av} versus the molecular weight (M_w). $K_{av} = (V_e - V_o)/(V_c - V_o)$, where V_e = elution volume, V_o = column void volume, V_c = geometric column volume. Straight line is the calibration curve calculated from the data for molecular weight standards (see Section 2.2.7 and Figure 2.1). Using the equation for the line, values for the molecular weights of CPG2_{CAT}, CPG2_{CAT'}, CPG2_{DIM} were calculated. Molecular weights of CPG2₂₃₋₄₁₅ are plotted for comparison.

6.5 Preparation of Isolated CPG2 Domains Leads to Greatly Improved NMR Spectra.

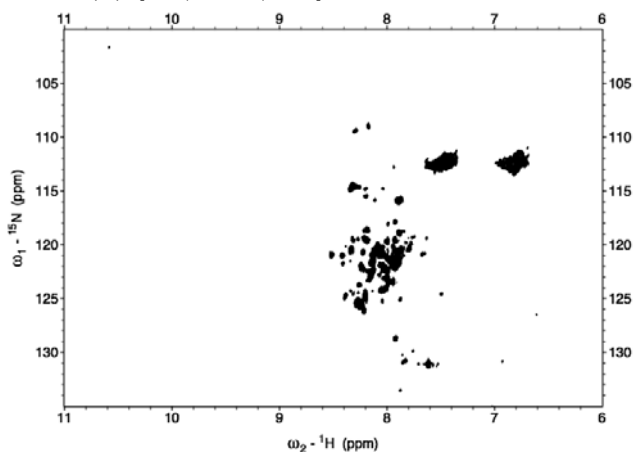
Two dimensional ^1H - ^{15}N HSQC NMR spectra were then acquired for all three protein constructs in Tris buffer (20 mM Tris, 100 mM NaCl, pH 7.3) to explore their suitability for further structural characterisation, and the spectra are shown in Figure 6.8. The 110-residue protein derived from the CPG2 dimerisation domain (CPG2_{DIM}), did not yield a well-dispersed ^1H - ^{15}N HSQC spectrum, with most of the backbone amide ^1H chemical shifts falling between 7.7 and 8.4 ppm (Figure 6.8-C). These results indicate that the dimerisation domain is highly unstructured in isolation and likely required the presence of the catalytic domain to fold correctly. The largest improvement in peak width was observed for the 176-residue CPG2_{CAT} protein, which produced a well-dispersed ^1H - ^{15}N HSQC spectrum suggesting a well-folded protein (Figure 6.8-B). However, given the low activity observed for this construct (3.7 U/mg, Table 6.1), it is unlikely to be carried forward for structural investigation in the future.



(a) $[U-^{13}/^{15}N]$ -labelled CPG2_{CAT}.



(b) $[U-^1/H^{15}N/^{13}C]$ -labelled CPG2_{CAT}.



(c) $[U-^{15}N]$ -labelled CPG2_{DIM}

Figure 6.8: 1H - ^{15}N HSQC spectra of the different domains.. Approximately 0.8 mM of A) CPG2_{CAT}, B) CPG2_{CAT}' and C) CPG2_{DIM} in 20 mM Tris and 100 mM NaCl pH 7.2 at 25°C on 700 MHz using trosyettf3gpsi and hsqcetf3gpsi pulse sequences for CPG2_{CAT}, CPG2_{CAT}' and CPG2_{DIM}, respectively.

Significantly improved chemical shift dispersion was observed for the CPG2_{CAT} protein (Figure 6.8-A) suggesting that the non-contiguous catalytic domain is able to fold in the absence of the dimerisation domain. However, an overlay of the CPG2_{CAT} HSQC spectrum with the corresponding spectrum of CPG2₂₃₋₄₁₅ indicated that many of the chemical shifts have changed upon removal of the dimerisation domain, likely because of the absence of contacts between these domains (Figure 6.9). Nevertheless, comparison of the peaks in the CPG2_{CAT} HSQC spectrum to those in the corresponding spectrum of CPG2₂₃₋₄₁₅ did show some improvement in line width and variable intensity suggesting improved solution behaviour for this smaller construct while retaining activity against methotrexate (Figure 6.5-C and Table 6.1).

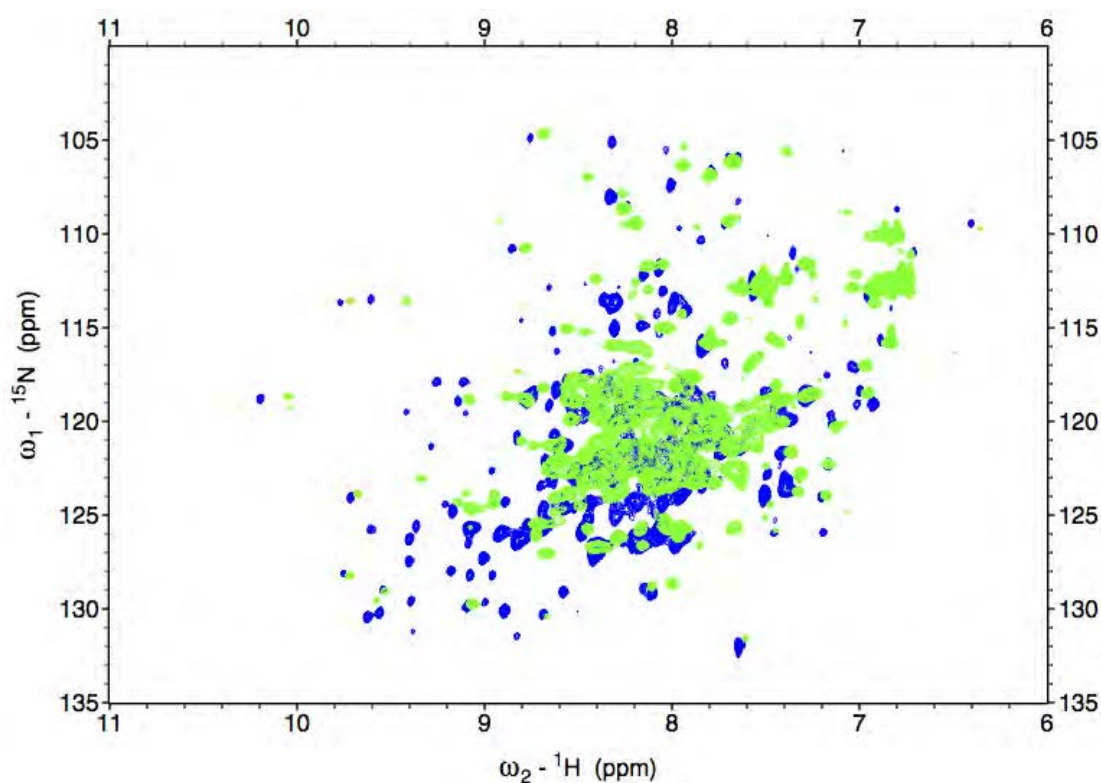


Figure 6.9: ^1H - ^{15}N HSQC spectra comparison of CPG2_{CAT} and CPG2₂₃₋₄₁₅. Overlay of CPG2₂₃₋₄₁₅ (in blue) with CPG2_{CAT} (light-green) spectra. Used 1 mM sample protein in 20 mM Tris and 100 mM NaCl, pH 7.2 at 25°C using troyetf3gpsi pulse sequence (700 MHz).

Since we were quite interested in the CPG2_{CAT} construct, we decided to do a rapid secondary structure determination to analyse how different the folding properties of the engineered catalytic domain were compared to the structure of the catalytic domain within the full-length CPG2 protein. We used CD to analyse how ligating these two non-contiguous primary sequences to form the catalytic domain affected the final conformation for the catalytic domain generated. DichroWeb, an online tool, was used to predict the secondary structure from the data obtained via circular dichroism (Whitmore and Wallace 2004; Whitmore and Wallace 2008).

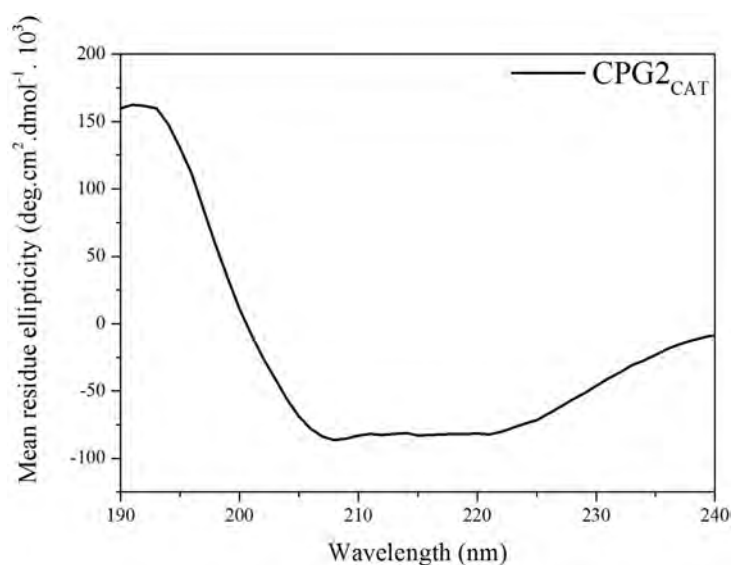


Figure 6.10: CD spectra of CPG2_{CAT}. According to DichroWeb the secondary structure content was 79% α -helix, 11% β -strand and 10% random coil (Whitmore and Wallace 2004; Whitmore and Wallace 2008).

6.6 Conclusion

A “divide-and-conquer” approach was utilised to study smaller domains of CPG2₂₃₋₄₁₅ with solution properties (e.g. faster tumbling, slower relaxation) more amenable to NMR measurements. Considering the molecular domain organisation of CPG2, we chose to divide the protein into three smaller fragments: the dimerisation domain fragment (CPG2_{DIM} residues 213-322); the full catalytic domain fragment (CPG2_{CAT} composed of residues 23-214 and residues 323-415 linked via a non-native alanine); and the largest contiguous region of the catalytic domain (CPG2_{CAT'} composed of residues 23-198) (Figures 6.1 and 6.2). CPG2_{DIM} did not yield a well-dispersed ¹H-¹⁵N HSQC spectrum suggesting that the dimerisation domain is highly unstructured in isolation and likely requires the presence of the catalytic domain to fold correctly (Figure 6.8-C). The methotrexate-based activity assay also revealed that CPG2_{CAT} generated a low activity, one comparable to full-length CPG2. On the other hand, as anticipated, CPG2_{CAT} was active, but less active than CPG2₂₃₋₄₁₅ suggesting that the dimerisation domain plays a role in catalysis (Table 6.1). Plus, significantly improved chemical shift dispersion was observed for the CPG2_{CAT} protein suggesting that the non-contiguous catalytic domain is able to fold in the absence of the dimerisation domain (Figures 6.8-A).

In an attempt to get a better-resolved spectra, we also tested a smaller construct, CPG2_{CAT'}. The 176-residue CPG2_{CAT'} showed the largest step-up in peak width and the protein produced a well-dispersed ¹H-¹⁵N HSQC spectrum suggesting a well-folded protein (Figure 6.8-B). However, given the low activity

observed for this construct (3.7 U/mg), it will not be carried forward for structural investigation in the future (Table 6.1). Most importantly, the results revealed the implication of the C-terminal region of the catalytic domain in substrate recognition / activity.

Chapter 7

Discussion and Perspectives

The three-dimensional structure of CPG2 is of great interest due to its therapeutic applications in the treatment of cancer and auto-immune diseases. Early X-ray crystallography investigation of CPG2 (Rowsell et al. 1997) demonstrated that it consists of two domains, the catalytic domain which delivers the substrates for the two zinc ions in the active site, and the dimerisation domain which is thought to stabilise the dimeric form of the enzyme (Figure 1.3). This crystallographic data helped to better inform early rational drug design efforts (Rowsell et al. 1997; Khan et al. 1999), but further attempts at co-crystallisation of CPG2 with ligands, substrates or inhibitors have not been reported and very little progress has been made since these early studies to establish key features of CPG2 such as the location of its active site, the presence of additional ligand-binding sites, its stability, its oligomeric state, and the molecular basis of activity. Thus far, our current working understanding of the molecular basis of CPG2 activity is largely based on sequence / structural

homology and molecular modelling (Lindner et al. 2003). Solution state nuclear magnetic resonance (NMR) spectroscopy is a unique tool for studying protein structure, stability, dynamics, and ligand binding properties that could inform structure-based drug discovery and provide a fundamental understanding of CPG2 function (Wuthrich 1986; Stark and Powers 2012; Clarkson and Campbell 2003). The work presented in this thesis sought to gain further understanding of the thermodynamic and structural aspects of the mode of binding of the inhibitors Mologic Ltd. (Bedford, UK) has developed to target CPG2 enzyme, but also the key features of CPG2, its stability and molecular basis of activity.

This chapter brings together the work pursued in order to address these questions, and outlines how the results of the work have advanced our knowledge.

7.1 First Reported Method for High-Yield Expression of Active and Soluble Mature CPG2₂₃₋₄₁₅.

Chapter 3 describes the first reported method for high-yield expression of active and soluble mature CPG2 (in the absence of the leader peptide) and its individual catalytic and dimerisation domains in *E. coli*. We have used this method to routinely produce milligrams quantities of ^{15}N and $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ isotopically-labelled protein suitable for NMR studies. Using a range of biophysical methods, the impact of solution conditions such as ionic strength, temperature and pH on the fold and stability of CPG2 was demonstrated, as

well as its activity against one of its primary substrates, methotrexate. Conditions were obtained (20 mM Tris, 0.1 M NaCl, pH 7.3) that enabled preparation of concentrated and stable CPG2 samples suitable for NMR studies.

7.2 The Aromatic Ring and Glutamate Groups are Essential Key Features in CPG2 Inhibitors.

Several attempts have been made to design potent inhibitors for CPG2. The first active inhibitors N-benzoylglutamate (Levy and Goldman 1967), which showed moderate inhibition activities against CPG2, and then thiocarbamate inhibitors (Khan et al. 1999), which unfortunately were found to be toxic against LS174T cells.

Recently a series of new inhibitors were synthesised by Mologic (Bedford, UK) highlighting important structural aspects present in substrates such as MTX for CPG2 enzyme (Turra et al. 2012). These were tested for their CPG2 inhibitory activity by Mologic using the octet RED system, which allows affinity analysis of biomolecular interactions. A series of peptide clearing agents consisting of a binding molecule which specifically targets and inhibits CPG2 activity at non-tumourous locations were analysed, where the most potent inhibitors were CP06 and CP39 with an IC_{50} of 88.5 μ M and 61 μ M (Figures 4.1 and 4.2 respectively). The CP06 and CP39 inhibitors were also galactosylated to allow and expedite clearance of the inhibited CPG2 enzyme

via the liver by binding to the asialoglyco-protein receptor expressed by hepatic cells (Figures 4.1—CP67 and 4.2—CP42, respectively). However, only CP67 (galactosylated CP06) showed significant inhibition with an IC_{50} of 93 μ M.

We wished to investigate the reason why CP06 and CP39 retained or lost inhibition strength upon galactosylation. The major structural differences between these two inhibitors are the galactosylated moiety and the type of aromatic ring used. CP39 having degraded, we were left with CP06, CP42 and CP67 for further investigations. The results revealed that glutamate is essential for recognition and / or binding as protons of the glutamate group showed saturation while none of the CP42 seemed to be close to the CPG2 binding site. This data could explain the absence of inhibition observed for CP42 using the octet RED system and the very high K_D values for aromatic protons exhibiting saturation. Furthermore, although inhibitory properties can be improved by replacing the naphthalene ring (CP06) with a benzyl ring (CP39), the change in the aromatic ring type together with the galactosylated group seemed to have an effect on the inhibitory response of these ligands. As a matter of fact, CP06 and CP67 inhibitors revealed that the naphthalene ring may play a significant role in molecular recognition by CPG2 (Figures 4.7 and 4.8, and Table 4.1). Despite the benzyl ring of CP42 showing proximity to CPG2 (Figures 4.9), the extracted CP42 K_D from the STD-NMR titration experiments revealed weaker binding than that of the naphthalene ring in CP67 (Table 4.1). A distinct difference between CP67 and CP42 is the galactosylation type and the length of the linker separating the galactosylated end and the binding molecule inhibiting CPG2 (Figure 4.2—CP42). The galactose groups

may be too close in space to the "warhead" in CP42. This absence of conformational flexibility may affect the adaptability of the ligand-binding site, and this may explain the absence of saturation of the glutamate and aromatic groups, and therefore the lack of inhibitory activity of CP42 towards CPG2. The absence and presence of saturation for the glutamate group and benzyl ring, respectively, suggest that CP42 does not efficiently sit within the CPG2 active-site, but that its benzyl ring still forms weak interactions with the CPG2 active site.

Site-directed mutagenesis of the hydrophobic pocket which was assumed to receive the aromatic moiety according to the X-ray crystallography structure and computational simulations showed that residues K209, S210 and T357 within the pocket P1 are essential for substrate recognition MTX (Chapter 6). Mutation of these polar (K208 and S210) and aromatic (T357) amino acid by alanine in CPG2 seem to have abolished electrostatic and / or π - π stacking interactions respectively with the pteridine ring of MTX, and further supports the importance of the aromatic moiety for CPG2 substrate recognition. Furthermore, none of the protons of the galactosylated domain in CP67 revealed proximity to CPG2 protein. This result indicated that, firstly, the galactosylated moiety was available for the hepatocyte receptors to bind to for blood clearance of the complex, and secondly the galactosylated moiety did not interfere with the CP67 domain responsible for its inhibitory activity and CPG2 recognition.

^1H - ^{15}H HSQC data was used to determine the binding affinity of CP06 and CP67 to CPG2. In general, ^1H - ^{15}H HSQC is used to locate the binding site of

the ligands on the target protein. However, ^1H - ^{15}H HSQC spectrum assignment of a large protein like CPG2 was not feasible due to line-broadening of the NMR signals, spectral crowding and frequency degeneracy as a result of a large number of expected NMR signals (theoretically 400 peaks for CPG2₂₃₋₄₁₅). Hence prior to focusing on assigning cross-correlation peaks observed in the ^1H - ^{15}H HSQC spectrum to each CPG2 nuclei, a very simple method was used to study the binding affinity of the inhibitors to CPG2: chemical shift perturbation (Chapter 4). By monitoring the chemical shift changes during the titration for CP06, one type of exchange regime was observed for the CP06-CPG2 complex when comparing the chemical shift of uncomplexed and complexed CPG2 signals. Continuous linear chemical shift changes were observed for some resonances, while other resonances displayed a non-linear chemical shift changes characteristic of a conformational transition in the ligand-bound CPG2 (Williamson 2013) (Figure 4.18). The calculation of the binding constants (K_D) for the few residues displaying chemical changes yielded to micromolar range values between 40-493 μM (Figure 4.19), very similar to binding constants derived from STD-NMR data for CP06 where the majority K_D were below 500 μM (Table 4.1). On the other hand, we observed an intermediate exchange regime for the CP67-CPG2 complex as resonances with the most pronounced chemical shift perturbation broadened and disappeared at various locations within the ^1H - ^{15}H HSQC spectra upon titration. As a result, binding K_D constant could not be obtained for CP67 (Figure 4.21-F). Severe line-broadening, and the cross-peaks' shape and pattern, in the HSQC spectrum suggested an increased effect of CP67 on the CPG2 chemical shifts, in

comparison to CP06, which could be interpreted as a further change in the tumbling rate of CPG2 upon binding of CP67 to CPG2. Since assignment of the 42 kDa CPG2 enzyme was not completed, co-crystallisation screening was also conducted for X-ray crystallography to locate the inhibitors binding site (Chapter 3). The screening yielded crystals that are awaiting analysis.

7.3 Insight into CPG2 Substrate Recognition.

CPG2 is an interesting protein to study due to its implication in anti-cancer and autoimmune therapies. Despite the X-ray crystal structure of CPG2 being determined by Rowsell et al. (1997) and its overall structural similarities with related AAP and GCPII proteins (Lindner et al. 2003), which had their mechanisms of catalysis proposed via the computational methods, the MTX binding mode hypothesis was mainly established by computational studies and homology structure studies (Turra et al. 2012). Turra et al. (2012) suggested that residues Ser210, Lys208 and Thr357 (located in pocket P1) may be involved in hydrogen bonding with the nitrogen atoms of the pteridine moiety of MTX, and residues Arg324 and Gly360 located in another adjacent pocket (P2) appeared to be hydrogen bonded to glutamate. Rowsell et al. (1997) showed that mutation of amino acid Arg324 to alanine had lower activity against MTX degradation. However, the implication of the other residues mentioned above were never proven experimentally, and the reaction mechanism of CPG2 remains non-elucidated. Hence, the work was sought to complement the information obtained via X-ray crystallography and computational

simulation of the CPG2-MTX complex using site-directed mutagenesis. Chapter 5 was aimed at characterising the contribution of the different residues in the catalytic site and substrate recognition are introduced in Figure 5.3.

This is the first time that 8 mutants of CPG2 have been expressed, and tested for activity against MTX. Only two mutations withing the active site, H385A and E200A, showed activity against MTX and inhibitory activity in the presence of CP06 and CP67 (Figure 7.1). This suggested that CP06 and CP67 were effectively competitive CPG2 inhibitors. Since these residues are zinc co-ligands, the data indicated that the mutant proteins may have compensated for their absence some other way. H385A protein showed an activity of 82.93 U/mL, whereas both wild-type and E200A proteins presented higher activities against MTX (96.7 and 63.70 U/mg respectively). The greater activity levels could be attributed to higher flexibility observed within the wild-type and E200A proteins by CD melts. On the other hand E175A and D141A mutant proteins showed residual activity, although they exist as very flexible folded proteins. This suggest that these protein residues which are zinc co-ligands may be directly involved in tuning the electrophilicity of the zinc ions and its reactivity (Ataie et al. 2008), explaining the low activities observed against MTX. K208A, S210A and T357A mutant proteins also showed residual activity against MTX suggesting direct implication in substrate recognition of P1, which could hold in the pteridine moiety of MTX but also the aromatic ring present in all three inhibitors investigated in the thesis.

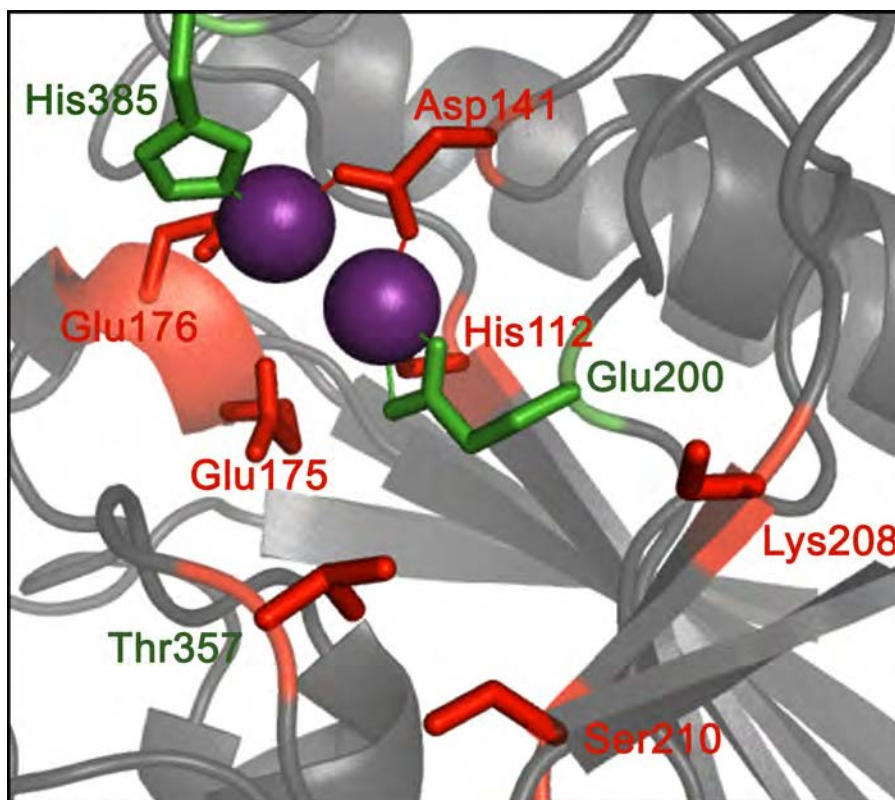


Figure 7.1: Summary of the mutations and their respective activities mapped onto the X-ray crystal structure of CPG2. The image shows the different residues, and their activity against MTX mapped onto the CPG2 binding site in the X-ray crystal structure (PDB: 1CG2, Rowsell et al. (1997)). Mutations coloured in red showed no activity, while mutations coloured in green demonstrated activity against MTX. [Pymol Viewer]

7.4 A "Divide-and-Conquer" Approach.

As opposed to crystallography, NMR could provide key information on the ligand-binding characteristics of CPG2, which is of high importance in applications using this enzyme (such as ADEPT). NMR can also be used to obtain information on conformational dynamics, which may have held back crystallographic approaches to ligand-binding studies thus far. Solution NMR studies must overcome a further limitation in that there is an upper limit for the molecular weight of proteins that can be readily analysed using this method, with few structural studies reported in the literature for proteins larger than thirty kilodaltons (Fernández and Wider 2003). Large, slowly tumbling proteins

in solution lead to a loss of resolution due to spectral crowding as well as broad NMR signals that are ultimately unobservable (Nietlispach et al. 1996). Although the NMR data described here for CPG2₂₃₋₄₁₅ were incomplete and thus did not yield resonance assignment, work is ongoing to use our optimised expression protocol and preliminary structural understanding to identify conditions that yield a full complement of 3D NMR data to facilitate assignment.

Hence a "divide-and-conquer" approach was used to study the smaller domains of CPG2₂₃₋₄₁₅ with solution properties (e.g. faster tumbling, slower relaxation) more applicable to NMR measurements. CPG2_{DIM} did not yield a well-dispersed ¹H-¹⁵N HSQC spectrum suggesting that the dimerisation domain is highly unstructured in isolation and likely requires the presence of the catalytic domain to fold correctly (Figure 6.8-C). It is the first time that the ability of the catalytic domain in CPG2 to fold independently without the dimerisation domain is reported and the components necessary for activity has been narrowed down. Specifically, the CPG2_{CAT} construct shows great promise for downstream NMR studies as it has favourable solution properties and retains key properties of the parent protein, namely enzymatic activity and the ability to self associate. Although the largest contiguous region of the catalytic domain (CPG2_{CAT'} composed of residues 23-198) showed the largest step-up in peak width and a well-dispersed ¹H-¹⁵N HSQC spectrum, it showed low activity. The results revealed that the C-terminal region of the catalytic domain is implicated in substrate recognition and activity.

7.5 Future Directions.

The first reported method for high-yield expression of active and soluble mature CPG2 (in the absence of the leader peptide) and its individual catalytic and dimerisation domains in *E. coli* was described. Such a method is of key importance for future structural, functional and binding studies involving this enzyme, which was hampered by difficulties in obtaining large yields of the protein as well as labelling of the protein. The NMR techniques useful to the investigation of different recognition processes have been successfully applied. In particular a versatile protocol based on STD-NMR and chemical shift perturbation experiments for the screening of putative inhibitors, towards the rational design and optimisation of novel CPG2 inhibitors have been provided. Moreover some more insights on the study of CPG2 binding requirements via site-directed mutagenesis and screening of putative inhibitors have been presented. This work establishes the support for future work.

To reinforce the NMR study conducted during this project, surface plasmon resonance (*SPR*) (Johnsson et al. 1991), isothermal titration calorimetry (*ITC*) (Ciulli et al. 2006; Torres et al. 2010), ligand-observed mass-spectrometry analysis (Chen et al. 2015) could also be employed. In SPR studies, proteins are immobilised on a metal laminated plate and ligands flow past. When a ligand binds to a protein, SPR spectroscopy quantifies using an optical device the changes in refractive and reflective properties of the metal surface, which are correlated to the mass of the ligand and the mass of the protein. The appreciable benefit of this technique is that only small amounts of

protein is required (unlike for NMR and ITC) and affinity can be acquired from the recorded SPR spectrum representing a constant flow system, where ligand saturate the protein and then wash off, and hence, k_{on} and k_{off} can be determined. SPR tolerates screening of several hundreds of compounds in a few days using the same SPR plate, and is well suited for prioritising subsequent X-ray co-crystallisation studies. This screening method could eventually be used to rapidly screen for potent inhibitors from the second generation of ligands produced in the near future by industrial.

In order to design specific inhibitors, structural information on substrate recognition with atomic resolution and on the catalytic mechanism is required. The advantage of crystallography is the more detailed info in well-defined parts of the molecule, and the absence of size limitation, which restricts solution NMR to molecule below ≈ 30 kDa. The next step in understanding the catalytic properties of CPG2 should be the establishment of a model describing the reaction mechanism of substrate hydrolysis by CPG2. This will necessitate the investigation of residue Gly360 located in P1 pocket that was not achieved, and analysis of potential co-crystals generated during this study (Chapter 3). Hence, solving the X-ray structures of the different mutant proteins could enable structural comparison of the substrate binding cavity, and possibly give insight in the CPG2 catalytic mechanism, which could further explain the activity of these mutant proteins, or lack of, against MTX.

Additionally, the CPG2₂₃₋₄₁₅ construct was crystallised without difficulty in this project and exhibited very similar structure to the full-length CPG2 published by Rowsell et al. (1997). Unfortunately, attempts at co-crystallisation

of CPG2₂₃₋₄₁₅ and its inhibitors, (i.e. CP06 and CP67 (Chapter 3) indicate that the protein packed in crystals may retain biological activity creating a challenge for the co-crystallisation process. In this case, we could potentially use inactive enzyme for co-crystallisation in order to dissect the molecular reaction mechanisms (i.e. an inactive CPG2₂₃₋₄₁₅ mutant created for the purpose of this project).

The previous rationale could be also adopted for protein-ligand studies using NMR spectroscopy. The inactive CPG2₂₃₋₄₁₅ construct, CPG2_{CAT'} of 20 kDa size and amenable for NMR studies (Chapter 6), could be used for determination of CPG2₂₃₋₄₁₅-inhibitors complexes. Using a relatively standard triple-resonance NMR experiments for standard backbone assignments of CPG2_{CAT'}, and chemical shift perturbations utilising HSQC titrations of [U-¹⁵N, ¹³C]-enriched CPG2_{CAT'} by CP06 and/or CP67 ligands, the resulting changes in the chemical shifts of the protein could be monitored, assigned to a specific protein residue and mapped onto the X-ray structure of the full length CPG2 by Rowsell et al. (1997) to localise the inhibitor binding site.

Furthermore, it was unambiguous from the outset that there would be several obstacles to overcome in order to use CPG2 in ADEPT. It was recognised that if foreign enzymes and those of bacterial origin, such as CPG2, presented the necessary specificity of action (Chapter 1), there would be an immunogenicity issue that would have to be discussed (Bagshawe 1989). Immunogenicity is the capability for an entity, with an antigen and an epitope, to induce a cell-mediated answer in the human body. To reduce immunogenicity in CPG2, a T-cell modification approach based on the HLA

binding, a prerequisite for T-cell stimulation, could remove antigenic eptitopes of CPG2 in order to weaken a durable and high-affinity immune response. Several immunogenic regions have been located using T-cell proliferation studies and *in silico* analysis (Chester et al. 2005). However, changes in amino acid sequence in enzyme can immediately lead to loss of catalytic activity. An imperative condition is that the modifications introduced do not interfere with structure, function or stability. In this thesis, the development of a less immunogenic CPG2_{CAT} protein (Chapter 6), was successfully demonstrated to be enzymatically active without the presence of the dimer subunit of CPG2. This shorter enzymatically active construct could be used as a foundation to produce less immunogenic CPG2 enzymes. Therefore, further work is necessary to reduce the immunogenicity barrier by further removal of the remaining problematic epitopes in CPG2_{CAT} protein to make ADEPT safer for patients.

Appendices

Appendix A

Study of Synthetic Peptides

Mimicking β 3-Loop Antibody

Binding Site in hCG and LH

Proteins by NMR Spectroscopy

A.1 Introduction.

Cancer treatments are more likely to work successfully if cancer is detected early. Today, most of the cancers are detected by biopsies, a painful procedure. An immunoassay provides a safer and non-invasive procedure. Tumour markers are molecules appearing in blood or tissues that are associated with cancer (Darwish 2006). The immunoassay approach allows quantification of a wide range of compounds including these biomarkers, which are indicators of cancer

prognosis. The aim of this side-project was to investigate the solution structure of polypeptides derived from a sequential epitope known as the $\beta 3$ -loop, which is present in all members of the glycoprotein hormone family. This chapter describes the different NMR experiments used to identify potential secondary structures within these small peptides and the obstacles associated with peptide analysis by NMR.

A.1.1 Structure Dependence in 8G5 Antibody Binding of hCG β_{66-80} Epitope.

Human chorionic gonadotropin (hCG) is used as a biomarker in pregnancy testing and diagnosis of hCG-secreting cancers, where tests are carried out as immunoassays. However, the hormones human chorionic gonadotropin (*hCG*), luteinizing hormone (*LH*), follicle-stimulating hormone (*FSH*), and thyroid-stimulating hormone (*TSH*) from the glycoprotein family consist of an identical α -subunit coupled to an hormone specific β -subunit. The α -subunit of the four hormones of the glycoprotein family are interchangeable and each is 92 amino acid residues in length. Despite also featuring substantial sequence homology, β -subunit contain the hormone-binding regions and influence the activity of each hormone (Pierce 1971). As a result, the β -subunit is targeted in all hCG immunoassays. Because of similarities in sequence and structure, the demand for specific immunoassays that are able to discriminate between these different proteins of the glycoprotein hormone family is required. Moreover, hCG is present in a wide-range of variants, diverging both in glycosylation state

and fragment sizes (O'Connor et al. 1994; Elliott et al. 1997; Birken et al. 1988). Hence, reliable hCG immunoassay should be able to detect all forms of hCG generated in cancer to the same extent. However, recent studies showed that care should be taken with sandwich assays, such as the sandwich ELISA assay. In a two-site immunoassay, hCG is sandwiched between two antibodies binding to different hCG epitopes, which could generate an increased number of false-negative results due to the doubled contingency of one of the antibodies not recognising certain hCG variants produced in later stages of cancer because of the absence of one of the indispensable epitopes (Griffey et al. 2013). Hence, focus has been on an immunoassay built around a well-known single hCG epitope site: the $\beta 3$ -loop from the β -hCG subunit shown in Figure A.1.

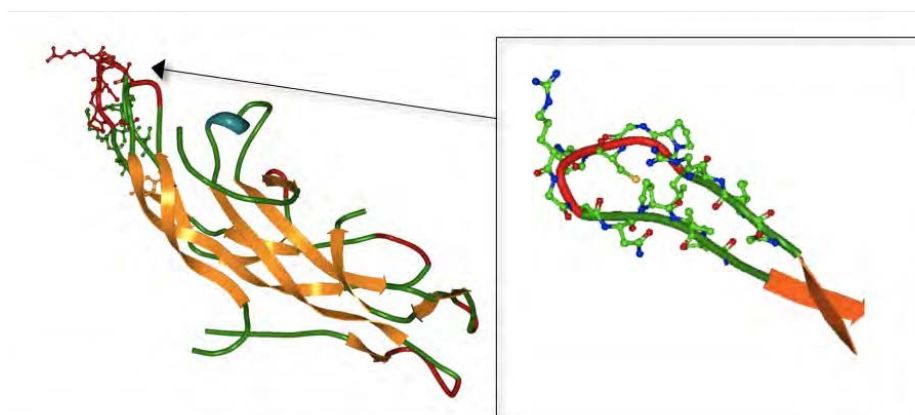


Figure A.1: Crystal structure of hCG. The full-length (*left panel*) and a close-up (*right panel*) of the hCG $\beta 3$ -loop as obtained from PDB ID: 1HRP (Lapthorn et al. 1994).

Gregor et al. presented binding affinity measurements demonstrating that 8G5 ($\beta 3$ -loop monoclonal antibody) was capable of distinguishing between hCG $\beta 3$ -loop (hCG β_{66-80}) and the closely related LH $\beta 3$ -loop (LH β_{86-100}). The $\beta 3$ -loop in hCG and LH differ only by one amino-acid; asparagine (Asn/N) in hCG and aspartate (Asp/D) in LH at position 77 and 87 respectively.

Postulating that the $\beta 3$ -loop contains self-contained secondary structural characteristics, Gregor et al. have synthesised hCG β_{66-80} as a synthetic peptide (SIRLPGCPRGVNPVV) and used point mutations in a range of peptide arrays to study whether a substitution at position 77, occupied by asparagine in hCG β , had an impact $\beta 3$ -loop-8G5 binding. The results from screening assays in Figure A.2 showed that the peptide mimicking the hCG $\beta 3$ -loop (shown in red) had a higher affinity for 8G5 antibody than the peptide derived from LH β_{86-100} (shown in green). Only one other peptide sequence containing histidine at position 77 (shown in blue), demonstrated binding to 8G5 antibody (60% that of hCG).

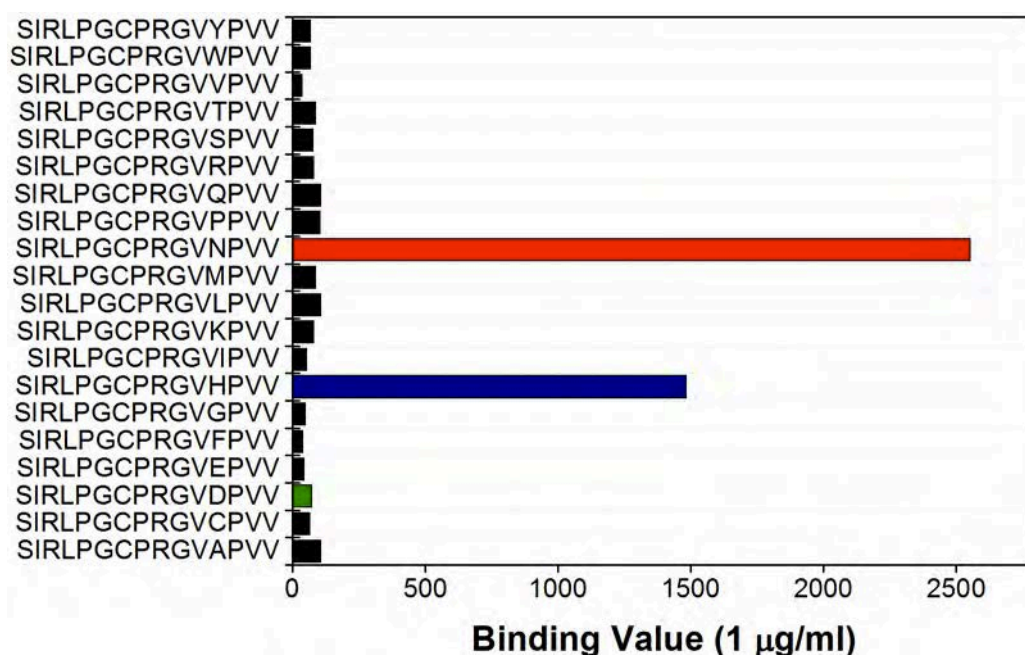


Figure A.2: Binding results from Pepscan - Mab 8G5. Binding result of the range of peptides arrays with a substitution at position 77, occupied by Asn in hCG β . SIRLPGCPRGVDPVV peptide equates to the $\beta 3$ -loop present in LH. SIRLPGCPRGVHPVV peptide is the only sequence, other than hCG $\beta 3$ -loop, exhibiting affinity to 8G5 antibody. Adapted from Gregor et al. (2011).

Additional computational studies suggested the presence of a putative hydrogen clasp between Asp77 and Val79 in hCG β_{66-80} may be crucial to

stabilise the β -turn observed over the P-R-G-V residues in hCG β_{66-80} . The presence of deprotonated carboxyl group in Asp97 of LH β_{86-100} may undermine the hydrogen clasp.

A.1.2 Project Aims.

These data implied that 8G5 was able to distinguish between hCG and LH β 3-loops due to the existence of well characterised structural difference between the hCG β_{66-80} and LH β_{86-100} peptides. The goal was to improve the understanding of antibody recognition of the β 3-loop epitope hCG β_{66-80} by studying the peptides shown in Table A.1, provided by Mologic and representing the various β 3-loops, using solution-state NMR spectroscopy. This could enable the development of more selective immunoassay methods for cancer detection and prognosis. New structural information suggested from NMR data, in combination with other techniques, could then be used to improve immunoassay methods, which is a valuable tool for cancer.

Table A.1: Sequence of peptides derived from hCG β 3-loop.

Name	Residue Number	Sequence	Origin
MOL174	66-80	SIRLPGCPRGVNPVV	hCG
MOL179	86-100	SIRLPGCPRGVDPVV	LH
MOL180	66-80	SIRLPGCPRGVHPVV	hCG with N77H mutation

A.2 Materials and Methods.

All NMR spectra were recorded on Bruker Avance 700 MHz NMR spectrometer using a 5-mm triple-resonance probe with 3-axis gradients at a temperature of 25°C. Spectra were calibrated with respect to residual water (4.719 ppm). Other spectrometer characteristics are reported in Section 2.2.10. All NMR spectra were processed using TOPSPIN 3.1 by Bruker (Coventry, UK) and analysed with NMRFAM-SPARKY (Lee et al. 2015).

A.2.1 Sample Preparation.

Sequences of the peptides used for the present studies are shown in Table A.1 of Chapter 7 on page 216. All peptides were synthesised using 9-fluorenylmethyloxycarbonyl (*Fmoc*) solid phase peptide synthesis, purified by high performance liquid chromatography, and confirmed by mass spectrometry by Mologic¹ (Bedford, UK).

For NMR experiments, the samples were prepared in 10% D₂O using 10 mM phosphate buffer and 1 mM DTT at pH 6.8. The peptide concentration was \approx 1.3 mM. In contrast with CPG2, these peptides are small and isotopic labelling was not required as they yield well-resolved homonuclear proton 2D-NMR spectra.

¹Dr James Schouten and Dr Joannah Towler (Mologic Ltd., Bedford, UK) produced the hCG/ β 3-loop peptides (MOL174, MOL179 and MOL180).

A.2.2 TOCSY.

TOCSY is the abbreviation for T^Otal C^Orelation S^Pectroscop^Y. It is also called HOHAHA (H^Omonuclear H^Artmann H^Ahn) (Braunschweiler and Ernst 1983; Bax and Davis 1985). The TOCSY experiment provides through bond proton-proton correlations due to scalar (i.e. through bond) couplings (Figure A.3-A). It provides through bond correlations between all the atoms in a single spin-system, i.e. all the protons in a single amino acid (Figure A.3-B). Each amino acid has a characteristic pattern of chemical shifts from which the amino acid can be identified (Figure A.3-C). However, some amino acids have identical spin systems and therefore identical signal patterns. They are: cysteine, aspartate, phenylalanine, histidine, asparagine, tryptophan and tyrosine ('AMX systems') on the one hand and glutamate, glutamine and methionine ('AM(PT)X systems') on the other hand. More or less, TOCSY constitutes a very efficient transfer element for the assignment of proteins, nucleic acids and small molecules (Wuthrich 1986).

A range of TOCSY experiments exist. In this work, TOCSY with a DIPSI-2 mixing sequence was used (dipsi2esgpph pulse sequence in Figure A.4) (Shaka et al. 1988; Hwang and Shaka 1995). After frequency labelling during t_1 , magnetisation is returned to the z-axis for isotropic mixing using the mixing sequence DIPSI-2, and then returned to the transverse plane for detection (Rance 1987). The TOCSY sequence works by putting in use a strong RF field, called a "spin-lock", along one axis. Under its influence the magnetisation is

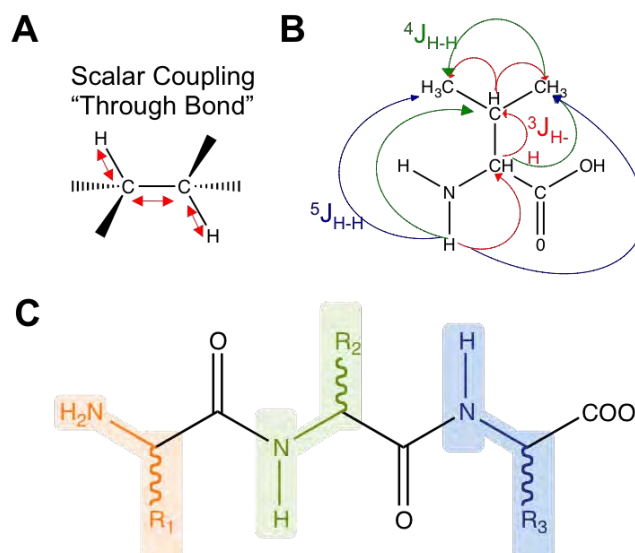


Figure A.3: Information provided by TOCSY experiment. A) TOCSY cross-peaks arise from scalar coupling. B) The number of transfer steps can be adjusted by changing the mixing time. Magnetisation can be transferred successively over up to 5 or 6 bonds as long as successive protons are coupled. C) TOCSY only allows identification of all protons within the same spin system, i.e. amino acid, because the presence of carbonyl groups disrupts the TOCSY transfer. Adapted from Roberts and Lian (2011).

transferred from one spin to another through the scalar coupling interactions. The efficiency of the magnetisation transfer depends on the J-coupling constant (Figure A.3-B), i.e. magnetisation is transferred more efficiently through large couplings, and is affected by the rate of relaxation during the isotropic mixing pulse. The presence of a cross-peak in TOCSY does not mean that the two spins are directly linked; instead, it suggests that the magnetisation can be transferred between the two spins through two or more bonds via scalar coupling between coupled spins (Figure A.3-B) (Cavanagh et al. 1995).

TOCSY spectra were recorded at 25°C, using a mixing time range of 50-130 ms. Data were acquired with 256 increments in the t_1 dimension and 2k increments in the t_2 dimension with 64 scans. The sweep width in both dimensions was 9803.92 Hz. Qsine window functions were used for processing in the F1 and

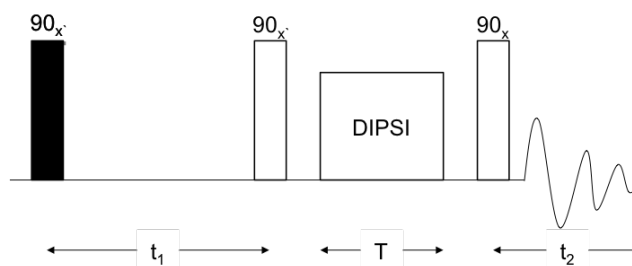


Figure A.4: General pulse sequence for TOCSY. In this sequence the isotropic mixing is carried out by mean of DIPSI. Adapted from Roberts and Lian (2011).

F2 dimensions, respectively.

A.2.3 NOESY.

NOESY stands for Nuclear Overhauser Enhancement Spectroscopy (Kumar et al. 1980). The NOESY experiment is crucial for the determination of peptide structure. It uses the dipolar interaction of spins (the nuclear Overhauser effect, NOE) for correlation of protons (Figure A.5-A). It identifies the chemical shifts of pairs of protons that are close together in space within the peptide structure. These protons may or may not be within the same amino acid. The correlation between two protons depends on the distance between them, but normally a signal is only observed if their distance is smaller than 5 Å. The NOESY will have both inter-residue correlations (Figure A.5-B, red arrows) as well as intra-residue correlations (Figure A.5-B, blue arrows), which allows the investigator to deduce which residue is next to which in the peptide chain. It correlates protons which are distant in the amino acid sequence, but also close in space due to tertiary structure. This is the most important information for the determination of protein structures.

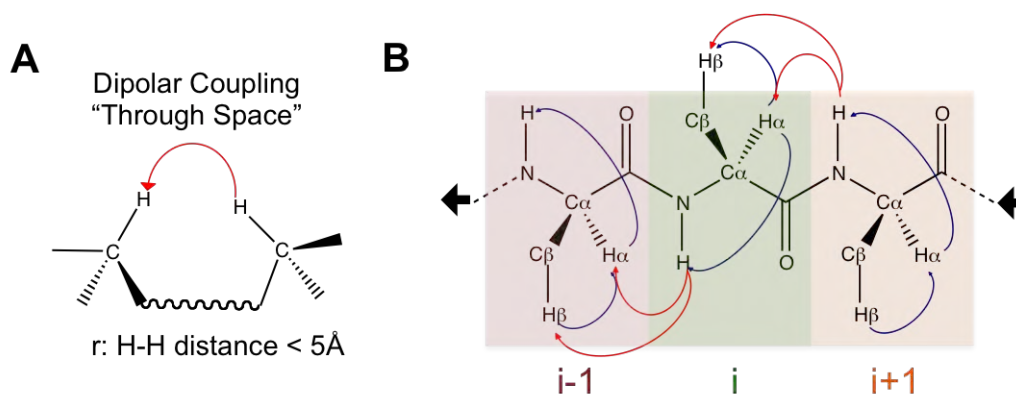


Figure A.5: Information provided by NOESY experiment. A) NOESY cross-peaks arise from dipolar coupling and is distance dependant, NOE cross-peaks only arise for protons within 5 Å of each other. B). NOESY provides intra-residue correlations (in blue) and also inter-residue correlations (in red), crucial for sequential assignment.

The NOESY pulse sequence scheme is shown in Figure A.6. The pulse sequence starts as usual with a 90° pulse followed by an evolution time t_1 . This delay is changed to provide chemical shift information in the F1 domain. Then another 90° pulse transmits some of the magnetisation to the z axis and during the following mixing period, the non-equilibrium z components will exchange magnetisation through the Nuclear Overhauser Effect (NOE). After some time (shorter than the relaxation time T_1), the transverse magnetisation is restored by the third pulse and detected. If relaxation exchange (or chemical exchange) has taken place during the mixing time, cross-peaks will be observed in the spectra.

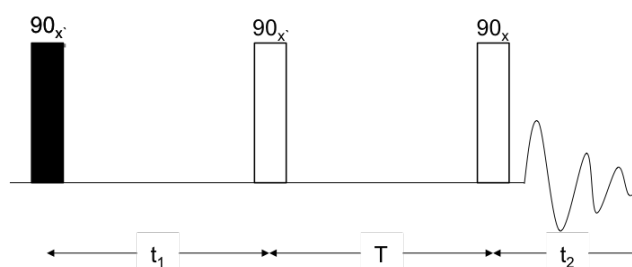


Figure A.6: General pulse sequence for NOESY. Adapted from Roberts and Lian (2011).

In this work, NOESY spectra using the noesyegpph pulse sequence were

recorded at 15°C, 25°C and 37°C at 600 MHz and 700 MHz (Hwang and Shaka 1995). The mixing time for peptide was varied from 60 to 500 ms. Typical two-dimensional data were acquired with 256 increments in the t_1 dimension and 2k increments in the t_2 dimension with 64 scans. The sweep width in both dimensions was 9803.92 Hz. Qsine window functions were used for processing in the F1 and F2 dimensions, respectively.

A.2.4 ROESY.

ROESY stands for Rotating frame Overhauser Effect Spectroscopy. The ROESY experiment is the rotating-frame analogue of NOESY, although there are a number of differences between the two experiments. ROESY uses a spinlock to align spins in the rotating frame. Cross relaxation happens during the spin lock time and develops at the rate comparable to T_1 (Figure A.7). The duration of the spin lock is set by the mixing time.

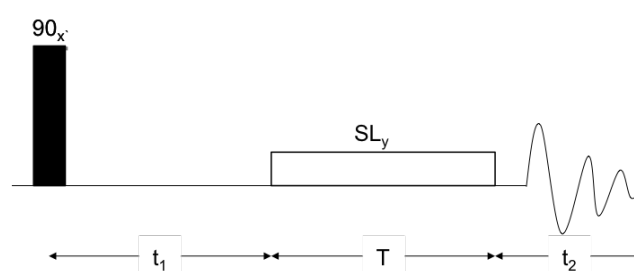


Figure A.7: General ROESY pulse sequence. This schematic shows the ROESY pulse sequence using continuous spin-locking. Adapted from Roberts and Lian (2011).

In this work, ROESY spectra using roesyegpph pulse sequence were recorded at 10°C, 25°C and 37°C (Hwang and Shaka 1995; Bax and Davis 1985). The mixing time for peptide was varied from 10 to 500 ms, similar to NOESY. Typical

two-dimensional data were acquired with 256 increments in the t_1 dimension and 2k increments in the t_2 dimension with 64 scans. The sweep width in both dimensions was 9803.92 Hz.

A.2.5 DQF-COSY.

DQF-Cosy stands for Double Quantum Filtered COSY. It was developed as an alternative to COSY, which gives information between two protons coupled via three bonds such as from a proton to its carbon, to the adjacent nitrogen, then to that nitrogen's proton (Figure A.8-A) (Rance et al. 1983; Piantini et al. 1982; Shaka et al. 1983). It contains the same information as does the COSY experiment. The value of 3J scalar coupling constant can be determined from the fine structure in the DQF-COSY cross-peak for spins without passive coupling partners, such as the $^3J_{\text{H}_\text{N}\text{H}_\alpha}$. Active couplings (those that lead to the cross peak) display anti-phase components (components with positive and negative signal intensity) (Figure A.8-B). The value of a three-bond J-coupling constant can be related to the intervening torsion angle, Φ , using the Karplus relationship (Figure A.8-C) (West et al. 1998; Pardi et al. 1984).

DQF-COSY spectra using the Bruker cosydfesgpph pulse sequence were recorded at 10°C, 25°C and 37°C (Hwang and Shaka 1995; Derome 2013). Typical two-dimensional data were acquired with 256 increments in the t_1 dimension and 2k increments in the t_2 dimension with 64 scans. The sweep width in both dimensions was 9803.92 Hz. The indirect dimension was zero

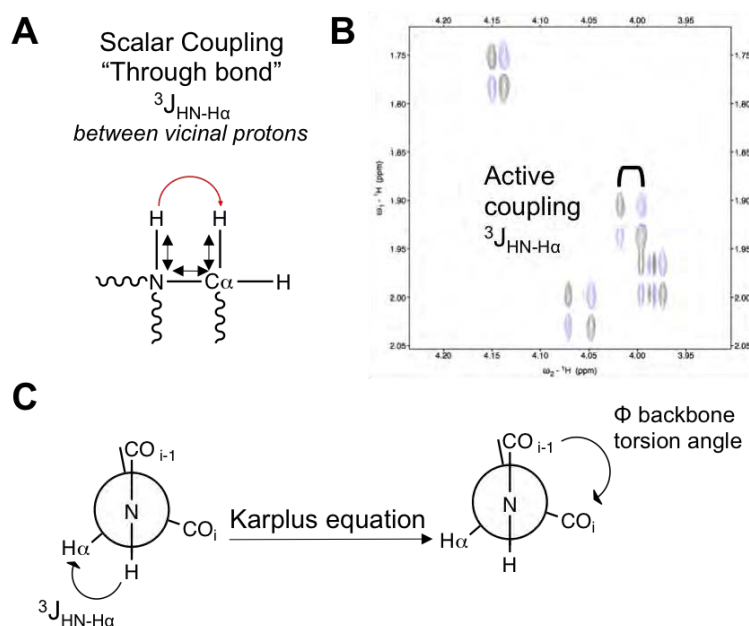


Figure A.8: Information provided by DQF-COSY experiment. A) COSY cross-peaks arise from scalar coupling. As opposed to TOCSY, magnetisation is only transferred up to 3 bonds. B) Cross-peaks are all antiphase and allow to obtain $^3J_{\text{HN-H}\alpha}$. C) Newman projections of peptide bonds that show how $^3J_{\text{HN-H}\alpha}$ is related to Φ backbone dihedral torsion angle via the Karplus equation.

filled during processing.

A.3 Discussion and Results.

A.3.1 ^1H NMR Chemical Shift Assignment of β 3-loop Peptides.

The chemical shift assignments of β 3-loop peptides were performed using TOCSY, NOESY and ROESY spectra obtained at 25°C by the sequential strategy proposed by Wuthrich (1986). For the purpose of describing the process, MOL174 has been used as an example (additional data in Appendix). For short small mixing periods, (e.g. 50 msec), TOCSY spectra gave 1-3 bond coupling. Correlation of more distant protons was only observable at higher

mixing time. With the mixing time set at 100 msec, correlations between H_N and protons down the side-chain of each amino acid were detectable. The extent of correlation depended mainly on the length of the mixing period. The “fingerprint region” of TOCSY was then used to establish each spin systems (Figure A.9-A). A single cross peak taken at the shift of one spin should in principle show the multiplets of all the spins that are part of the network of coupling to which the spin belongs (Figure A.9-B).

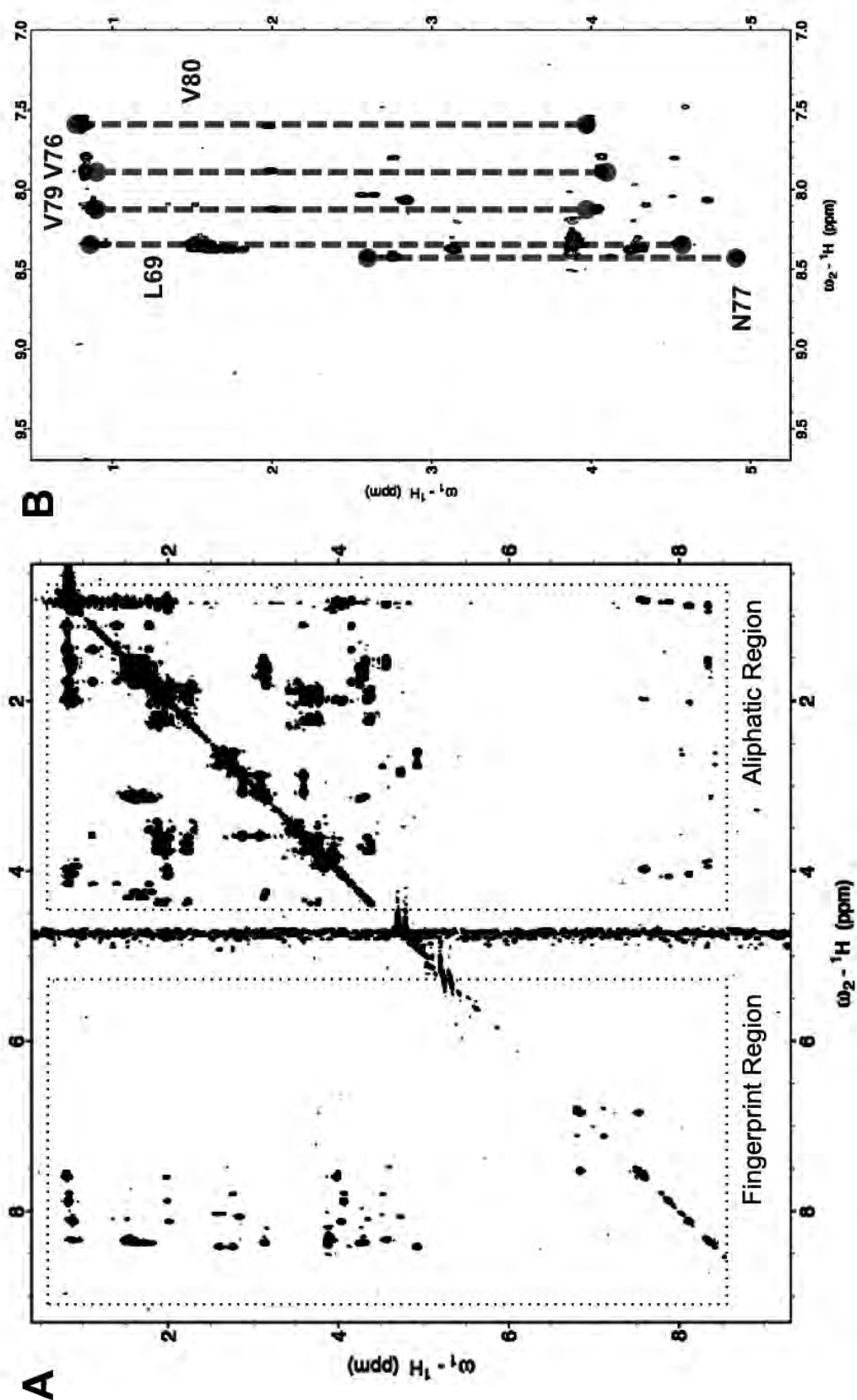


Figure A.9: Different regions of the 2D TOCSY spectrum. A) Full TOCSY spectrum and its different regions. "The fingerprint region" is located downfield with respect to the water signal, whereas the aliphatic is upfield. B) Spin systems assigned in the fingerprint region.

Many spin systems were determined by identifying as many through bond connections obtained from TOCSY spectra (Figure A.10). The H_N of isoleucine 2 was not observable in some peptides (i.e. MOL174), possibly due to exchange with the solvent, but the rest of the spin systems were visible in the fingerprint region (Figure A.10-A). Due to lack of H_N spin system in the TOCSY fingerprint region (Figure A.10-A). Due to lack of H_N spin system in the TOCSY fingerprint region, proline residues were only identifiable in the aliphatic region (Figure A.10-B). Some of the proline H_α cross-peaks, such as P73 of MOL174, P93 of MOL180 and P70 of MOL179, were not observable because their H_α chemical shifts were very close to the solvent suppression water signal, and therefore were attenuated by the water pre-saturation scheme used in the pulse sequence. The chemical shift assignment tables for all three peptides can be found in the Appendix.

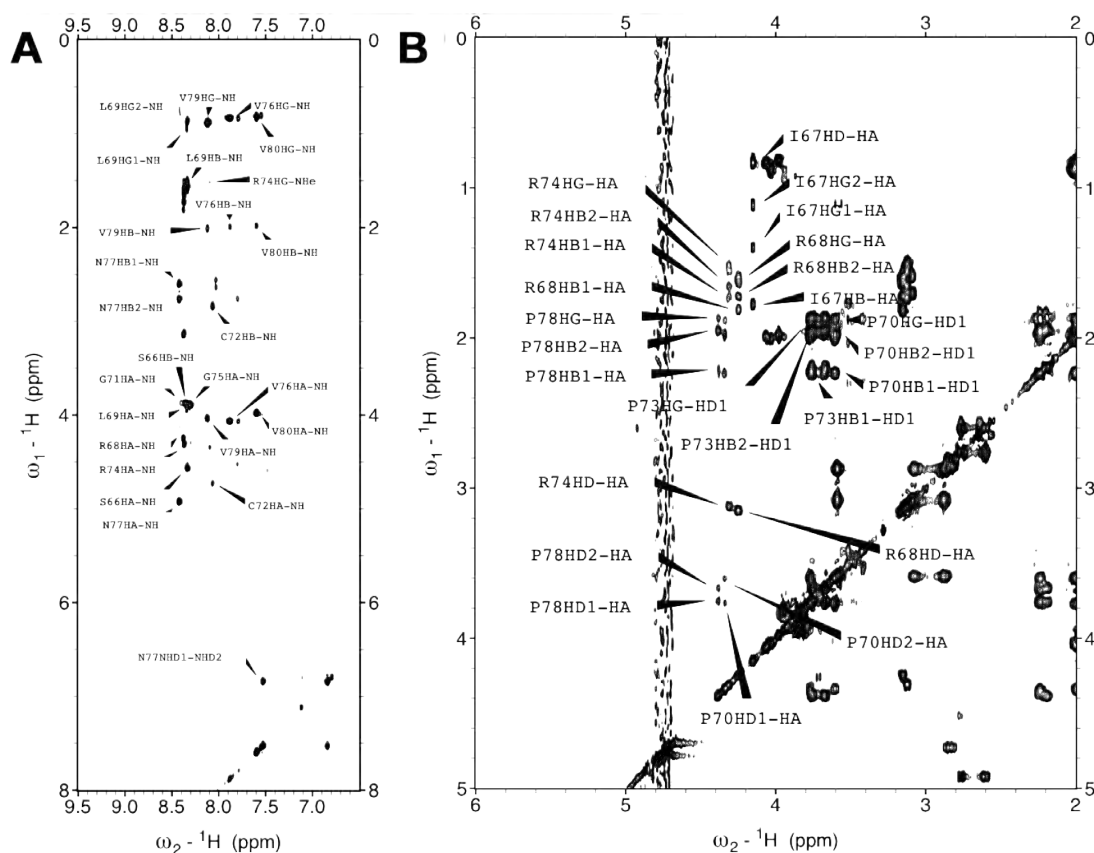


Figure A.10: 700 MHz 2D TOCSY spectrum of MOL174. A) Full TOCSY spectrum of MOL174 a mixing time of 100 ms. A) Fully assigned fingerprint region of the TOCSY spectrum of MOL174. B) Assigned aliphatic region of the TOCSY spectrum of MOL174.

A.3.2 Sequential Assignment of MOL174, MOL179 and MOL180 peptides.

Since there were redundant amino acid types in the sequence of each peptide, sequential assignment was only possible using the sequential connectivities observed through NOE or ROEs. All peptides showed very poor NOESY spectra, despite the fact that a wide range of mixing times were used. The NOE is created by the relaxation of one spin by a neighbouring spin. This event is proportional to the internuclear distance r , and the NOE is proportional to r^{-6} and time dependant. Figure A.11 shows that the NOE first builds-up (at cross-relaxation rates — relaxation mechanisms involving three spins) and then decays away, at roughly the spin-lattice relaxation rate ($1/T_1$) (Williamson 2009). Hence as the size of the compound studied increases, the cross-relaxation rate increases, and so does the decay rate (Williamson 2009). Thus, for big proteins, NOE reaches a maximum at low mixing times (around 300 ms), while for small peptides the maximum NOE can occur after several seconds.

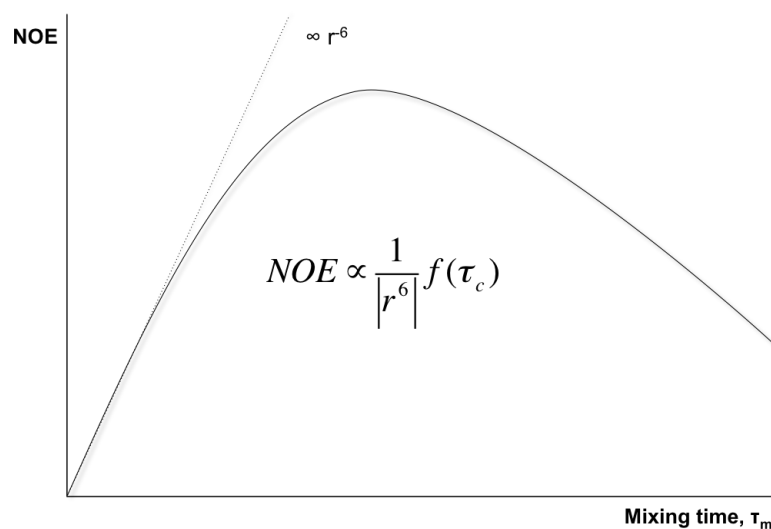


Figure A.11: Time dependence of NOE. The graph shows that at low mixing-time, the NOE quickly builds up, then starts to decay due to relaxation during the mixing time. The point of maximum NOE is shifted to higher mixing times for small peptides in contrast to huge proteins. Adapted from Claridge (2008).

Different NOE mixing times (60-500 ms) at 700 MHz were tested to check for the presence of NOE in the spectra. None of the mixing times tested enabled the observation of NOE peaks in the trio of peptides investigated. Figure A.12 shows the NOESY spectra at different NOESY mixing times ranging from 60 to 500 ms. Some H_N - H_α peaks were observable, but they corresponded to intra-residue NOE cross-peaks. We are only showing data for MOL174, but the other peptides behaved the same.

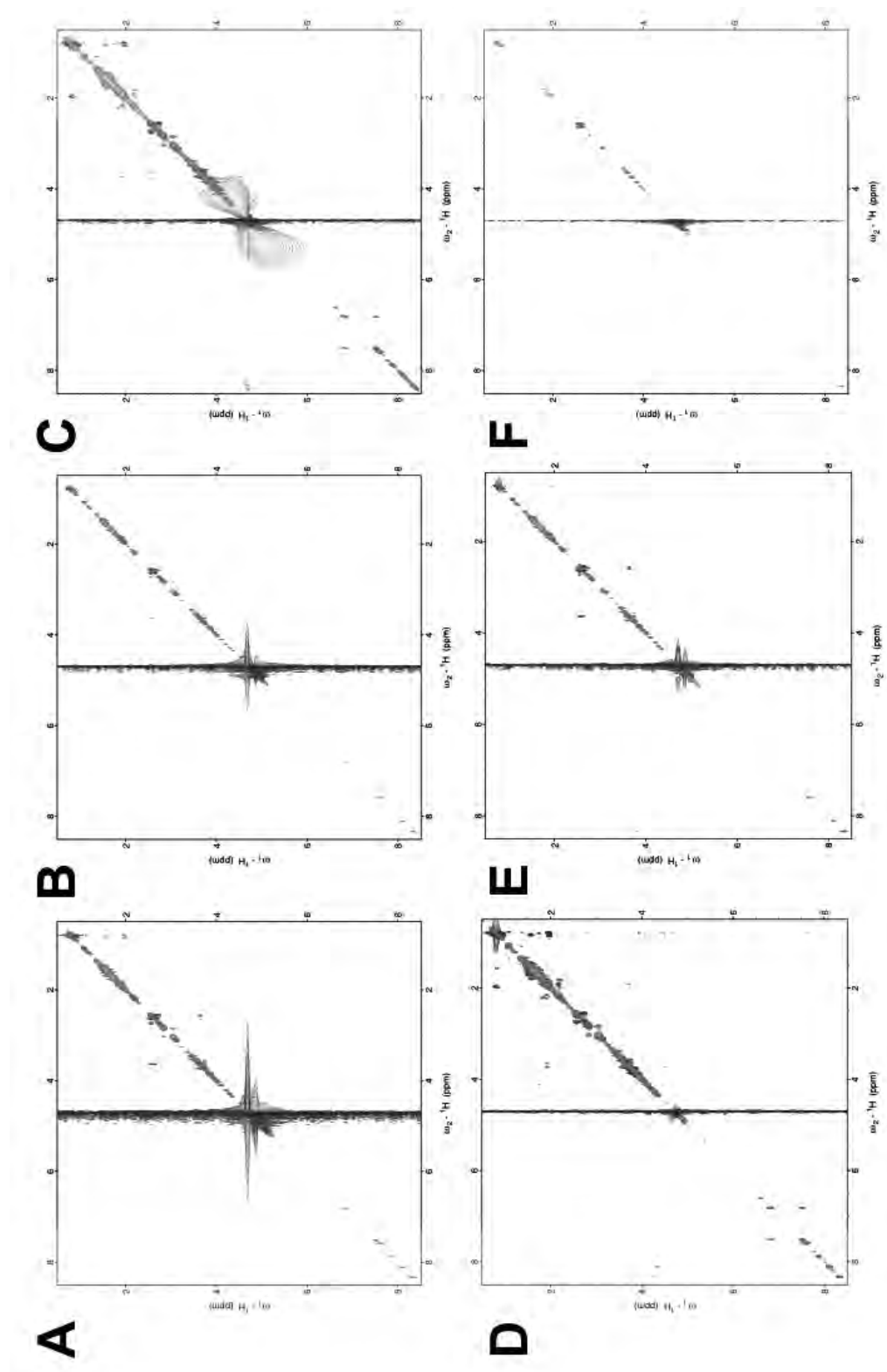


Figure A.12: 2D ^1H NOESY spectra of MOL174 at different mixing times. The NOESY spectra were collected at 25°C with a mixing time of *A*) 60 ms, *B*) 100ms, *C*) 200 ms, *D*) 300 ms, *E*) 400 ms and *F*) 500 ms.

There could be two reasons for these results: one is that the tumbling rate of the peptides generated NOEs with a near zero value intensity; the other reason may be due to high flexibility exhibited by the peptides in solution, thus not showing any NOEs. Figure A.13 shows the variation of the proton-proton NOE as a function of the molecular tumbling rate (Claridge 2008). The sign of the NOE depends on the magnitude of the correlation time and hence on the molecular weight. This curve represents the maximum proton-proton NOE as a function of molecular tumbling rates as defined by $\omega_0\tau_c$, where ω_0 is the spectrometer observation frequency and τ_c the correlation time — the time it takes for the compound to rotate through 1 radian (Wuthrich 1986). The curve has three distinct regions, which we shall refer to as fast, intermediate and slow motion regimes. For molecules that tumble rapidly in solution (short τ_c , hence low molecular weight molecule) the NOE has a maximum possible value of +0.5 or 50%. At the other end of the spectrum, molecules that tumble very slowly in solution (e.g. molecules with high molecular weight) experience negative NOE. In this motion regime, the maximum NOE of -1 or 100% can be obtained. Unfortunately, between these two extremes, at certain combinations of ω and τ_c the NOE changes sign and even becomes zero (at $\omega_0\tau_c \approx 1$). The tumbling time is such that the NOE intensity approaches zero. Unfortunately, the $\approx 1,500$ Da size peptides fell into this troublesome regime. To stay away from this regime, one could use a different magnetic field strength (ω_0), or alter the correlation time (τ_c) and the overall molecular tumbling rate by changing the temperature. Hence temperatures as low as 10°C and as high as 37°C were investigated, but no NOEs were observed. A lower magnetic field of 600 MHz was also tested and

was unsuccessful.

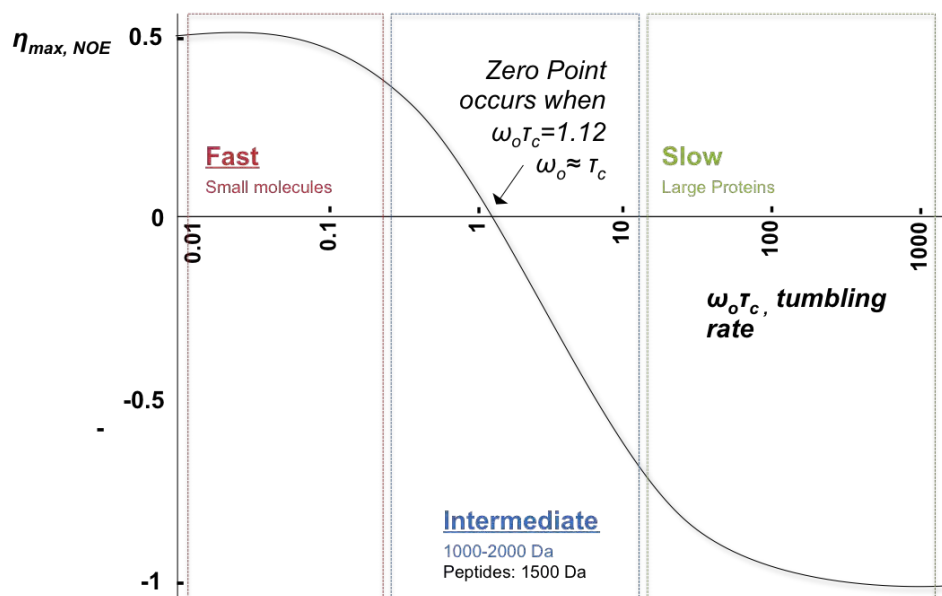


Figure A.13: NOE as a function of molecular tumbling rate. Adapted from Claridge (2008).

The only alternative left was to record a ROESY experiment instead of a NOESY (Bothner-By et al. 1984, Bax and Davis 1985). During the mixing time in NOESY, the magnetisation is aligned along the transverse plane of the static magnetic field, but in ROESY the magnetisation is aligned with a much smaller magnetic field, one that is created by a train of spinlock pulses and is oriented perpendicular to the static magnetic field. The result is that ROEs (NOEs in a rotating frame) in a ROESY experiment are always positive regardless of the correlation time, which is especially useful for intermediate molecular weight molecules, such as the peptides, where NOESY does not work (Keeler 2011) (Figure A.14-A). Unfortunately, the ROESY pulse sequence is almost identical to that used for TOCSY (Cavanagh et al. 1995). A possible complication is the relay of ROE through TOCSY resulting in false ROESY cross peak, as the opposite is also possible (Figure A.14-B,C,D). So care has to be taken when

running and analysing a ROESY experiment.

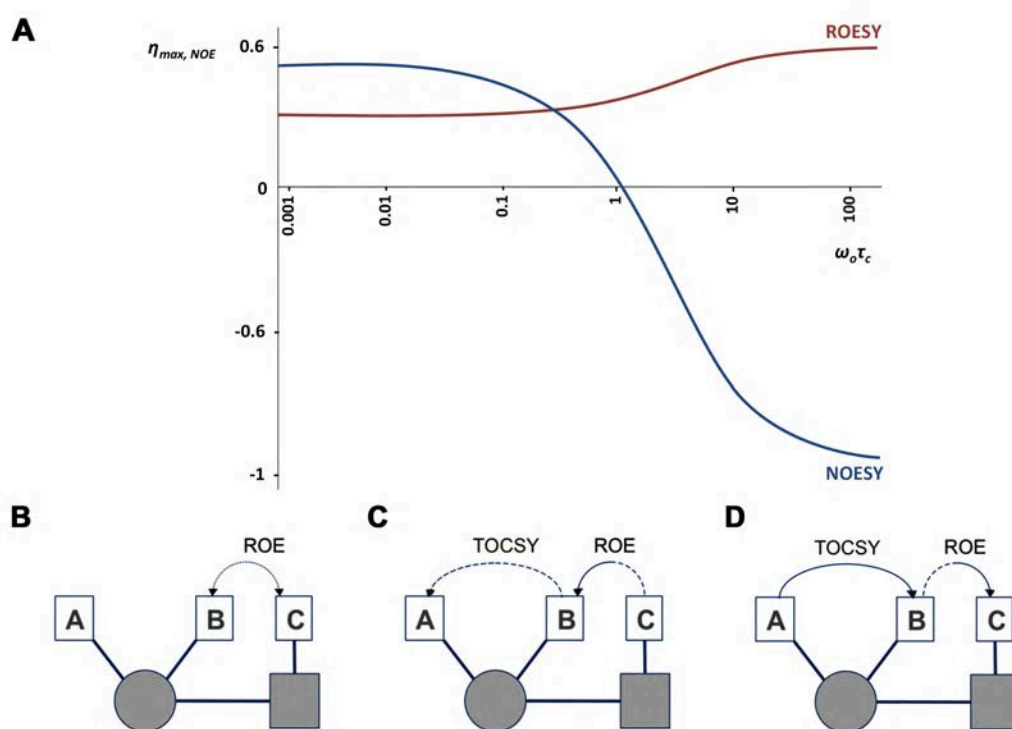


Figure A.14: Variation of the NOE sign in the NOESY and ROESY experiments, and ROE artifacts. A) The ROE is always positive, whereas the NOE goes through zero at a certain tumbling rate and becomes negative. B) Normal transfer pathway for an ROE. C-D) ROE artifacts due to the presence of spinlock pulses in the ROESY pulse sequence. Adapted from Claridge 2008.

These experiments resulted in the presence of ROE peaks for all three peptides at a mixing time of 200 ms (Figure A.15). Analysis of ROESY data revealed that MOL179 (D77) displayed very few ROEs compared to MOL174 (N77) and MOL180 (H77), which was indicative of a greater flexibility of MOL179 peptide. Moreover, only MOL174 featured long-range ROEs. This ROE was seen between $NH\epsilon$ of Arg74 and H_α of Val79. One of the residues is involved in the potential β -turn observed over the P-R-G-V sequence featured in the X-ray crystallographic structure of hCG β . On the other hand, the lack of these long-range ROEs within this region of the sequence for MOL179 and MOL180 may suggest the lack of stability of this turn in that part of the

peptide sequence.

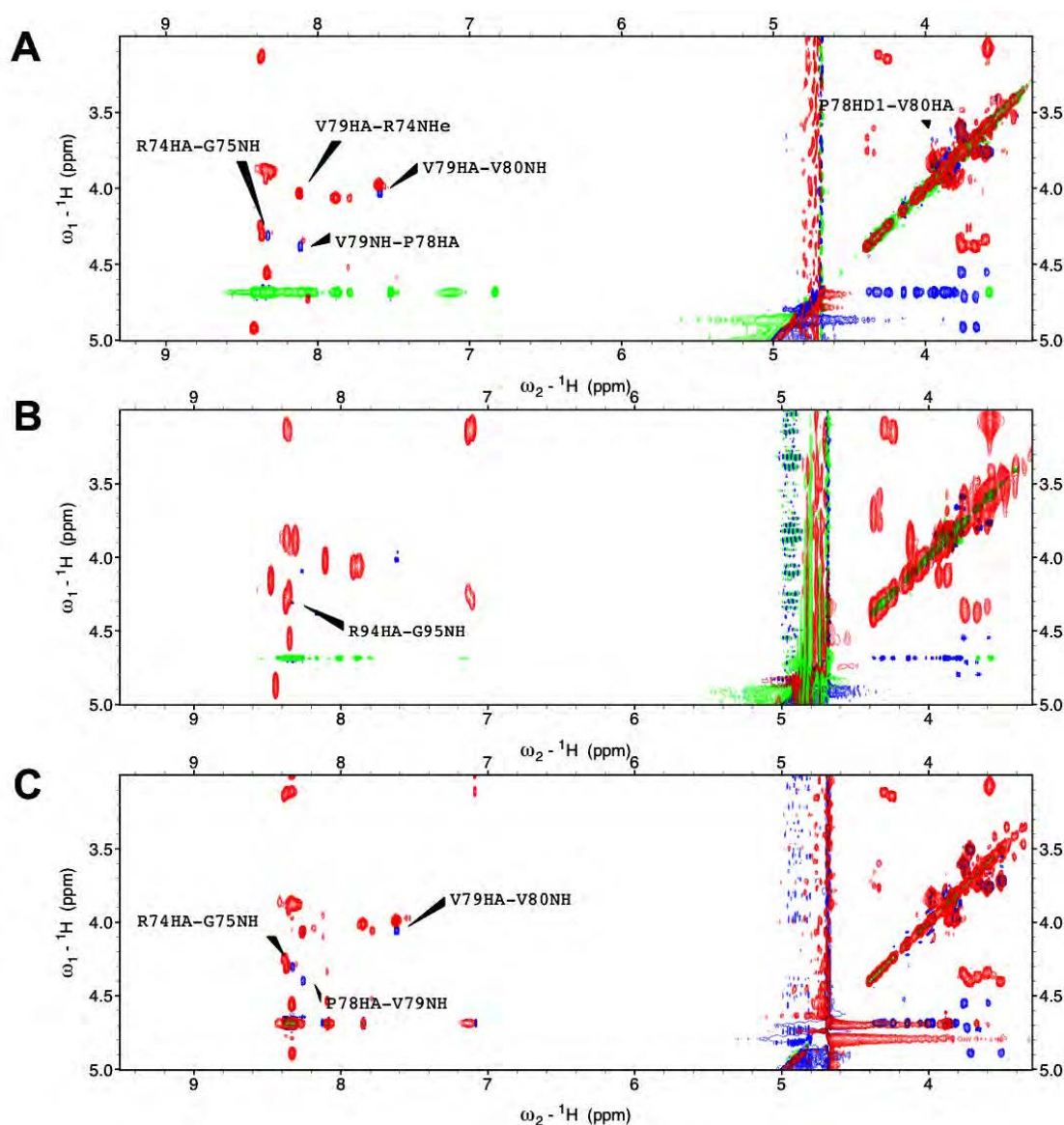


Figure A.15: Overlay of 2D ^1H ROESY and 2D ^1H TOCSY spectra in the amide region. TOCSY positive cross-peaks are coloured in red, while positive ROESY cross-peaks are coloured in blue. The ROESY spectra display the ROEs present in *A*) MOL174 (hCG β 3-loop with N77) *B*) MOL179 (LH β 3-loop with D77) and *C*) MOL180 (hCG β 3-loop with N77H mutation).

A.3.3 Secondary Structure Prediction using Chemical Shift Index.

The assignment of H_α allows the prediction of the secondary structure components present in the different peptides to be created utilising the chemical shift index (*CSI*) method introduced by Wishart (Wishart et al. 1991; Wishart et al. 1995). This is based on comparing the experimentally-obtained chemical shift values to a set of averaged random coil chemical shift values based on an array of seventy proteins whose structures are known (Wishart et al. 1991; Wishart et al. 1995). There is a strong correlation between H_α chemical shift and secondary structure as it can provide information about the local structure of a peptide. It has been shown that an upfield shift (e.g. towards lower ppm) from random coil values is observed with H_α chemical shift in α -helical like (and β -turn like) environment, while a downfield shift (e.g. towards higher ppm) is experienced when H_α is located in a β -sheet like region (Wishart et al. 1991). In this method three indicators are assigned, -1, 0 and 1, depending on the deviation from the random coil values. When an upfield shift is observed, -1 is attributed, reciprocally a downfield shift is tagged with a 1 index, while a shift close to random coil value is allocated a 0 indicator. Any sequence displaying a group of 4 or more -1 indices, not necessarily consecutive, and not disrupted by a +1 describes a helical (or β -turn) region, while any set of three or more successive +1 indices is recognised as β -sheet. The calculated chemical shift index for all peptides are displayed in Figure A.16. By the rules described above, MOL179 (D77) displayed a fairly unstructured secondary structure. MOL174 (N77) displayed β -turn like structure

in the R-L-P-R span, while MOL180 (H77) displayed a β -turn along a longer sequence span of P-G-C-P-R-G. MOL174 and MOL180 were the only peptides displaying a secondary structure and a binding affinity to 8G5, but with different strength (Figure A.2). Hence, these data show that a single point mutation seem to indeed disrupt the local structure of the peptide. Although the CSI showed different patterns for MOL174 and MOL180, the mutation from Asn to His at position 77 seemed to conserve the initial β -turn like conformation, while a mutation to Asp exposed a totally different CSI signature characteristic of an unstructured peptide conformation in solution.

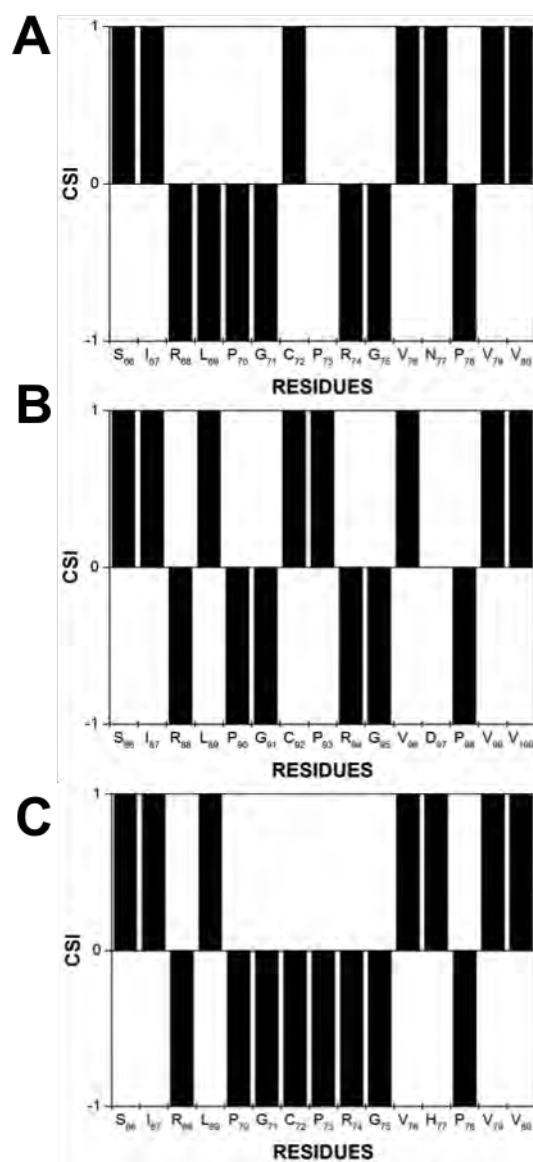


Figure A.16: CSI values of H_{α} . Calculated CSI values of C_{α} protons of A) MOL174, B) MOL179 and C) MOL180 peptides.

A.3.4 Secondary Structure Prediction using $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ Coupling Constants.

In addition to NOEs, $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ coupling constants were used to investigate the presence of a β -turn in all 3 peptides (Pardi et al. 1984). The $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ coupling constants were measured from the DQF-COSY spectra using the method described by Wüthrich, where the coupling constants were estimated from the 1D ^1H NMR spectrum as described in Figure A.17 (Pardi et al. 1984).

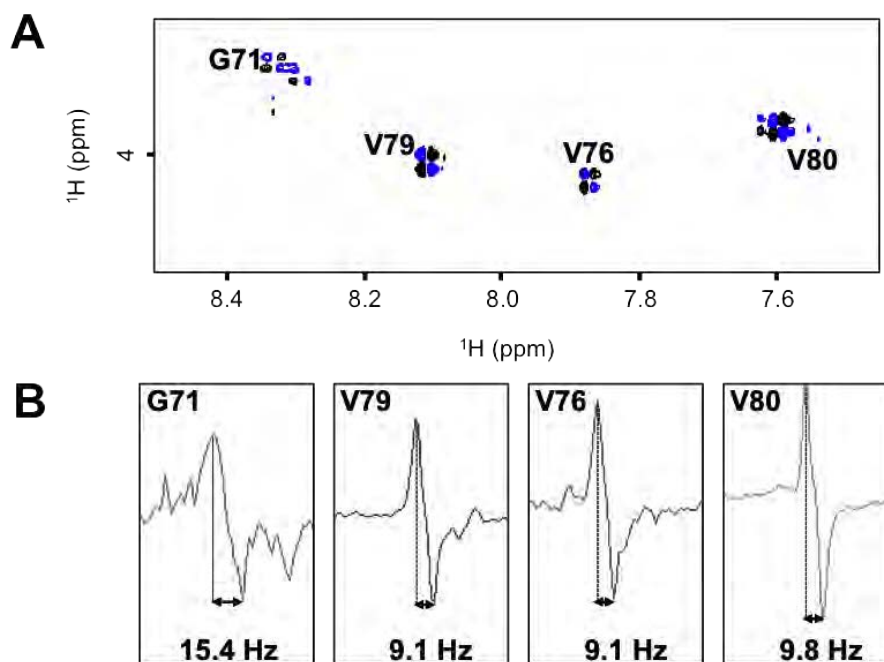


Figure A.17: H_N - H_α cross-peaks and cross sections parallel to F2 frequency axis from the DQF-COSY spectrum of MOL174. A) Cross-peaks of residues G71, V76, V79 and V80 in the DQF-COSY spectrum. B) Cross sections through the cross-peaks are shown. Measurement of the distance between the two peaks in each cross section (arrow) yielded to $^3J_{\text{H}_\text{N}\text{H}_\alpha}$.

However, the $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ -coupling constants should only be used in a qualitative manner and relative way in NMR analysis of peptides. This is because not only of the difficulty in measuring and interpreting $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ -coupling constants from DQF-

COSY data and conformational averaging. The Karplus curve derived from the Karplus relation (A.1) can be used to obtain the torsion angles of any ${}^3J_{\text{H-H}}$ -coupling constant between vicinal protons present in the peptide (Pardi et al. 1984).

$${}^3J_{\text{H}_\text{N}\text{H}_\alpha} = (6.4 \cos^2 \theta - 1.4 \cos \theta + 1.9) \text{Hz}, \theta = |\Phi - 60^\circ| \quad (\text{A.1})$$

The three-bond coupling constant between the intra-residual alpha and amide protons (${}^3J_{\text{H}_\text{N}\text{H}_\alpha}$) is the most useful for secondary structure determination as it can be directly related to the backbone dihedral angle Φ using Equation (A.1). For protons, three-bond J -couplings (3J) are typically around 7 Hz (Pardi et al. 1984). This number will vary immensely, depending on the local molecular structure of the peptide such as bond angles and torsional angles. Helical and extended conformations have very different values for Φ (-60° and -180° , respectively) which result in differences in ${}^3J_{\text{H}_\text{N}\text{H}_\alpha}$ (Richardson 1981). The calculated ${}^3J_{\text{H}_\text{N}\text{H}_\alpha}$ values shown in Table A.2 were used to investigate the secondary structure. The corresponding Φ dihedral angle values were extracted using Equation (A.1) (Pardi et al. 1984). Equation (A.1) predicts a maximum of 9.7 Hz. The presence of ${}^3J_{\text{H}_\text{N}\text{H}_\alpha}$ values above 9.7 Hz were observed, hence a Φ dihedral angle could not be generated from the Karplus equation for these coupling constants. These high ${}^3J_{\text{H}_\text{N}\text{H}_\alpha}$ -coupling constants values were observed in protein NMR possibly due to non-planar amide bonds (Demarco et al. 1978). The high values obtained for glycine residues suggested that peptides were quite flexible. Between Val76/96 and Val79/89 residues, the values started to range between 4.2 and 9.8 Hz. Experimental Φ values for ${}^3J_{\text{H}_\text{N}\text{H}_\alpha}$ of 4.2 observed in

MOL179 is -120° , which is typical of a helix structure. On the other hand $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ of ≈ 8 Hz ($\Phi = -151^\circ$) in MOL174 and MOL180 indicated the presence of β -sheet secondary structure (Pardi et al. 1984). These results suggested that there is a structural difference between the MOL174 and MOL180 peptide pair and MOL179 peptide in the region close to the P-R-G-V sequence, where the indispensable β -turn for 8G5 recognition is located.

Table A.2: ^3J -couplings between N and C_α protons, and their respective Φ dihedral angle in all three peptides. All the ^3J -coupling constants that are not present in an amino acid (i.e. proline) are left blank and missing with a "-". The dihedral angles that could not be calculated when ^3J -coupling constants were above 9.8 Hz and were marked with a "-".

Residue	$^3J_{\text{H}_\text{N}\text{H}_\alpha}(\text{Hz})/\Phi(^{\circ})$	$^3J_{\text{H}_\text{N}\text{H}_\alpha}(\text{Hz})/\Phi(^{\circ})$	$^3J_{\text{H}_\text{N}\text{H}_\alpha}(\text{Hz})/\Phi(^{\circ})$
	MOL174	MOL179	MOL180
Ser66/86	7.7/-148	9.8/-180	10.5/-
Ile66/87	-/-	-/-	-/-
Arg68/88	-/-	-/-	11.2/-
Leu69/89	9.1/-163	7/-142	9.1/-163
Pro70/90			
Gly71/91	15.4/-	7.7	25.9
Cys72/92	6.3/-137	-/-	7.7/-148
Pro73/93			
Arg74/94	-	16.1/-	9.8/-180
Gly75/95	15.4/-	16.1/-	13.3/-
Val76/96	9.1/-163	4.2/-120	9.8/-180
Asn/Asp/His	7.7/-148	6.3/-137	7/-142
Pro78/98			
Val79/99	9.1/-163	9.1/-163	8.4/-155
Val80/100	9.8/-180	9.1/-163	11.9/-

A.4 Conclusion.

Instead of obtaining a time-consuming full three-dimensional NMR structure of the protein of interest, the chapter showed that there are alternative and faster approaches to investigate structural features. Although the information obtained from three-dimensional protein structure is undeniable, the investigation of NMR chemical shift allowed to collect information about the structure of the β 3-loop, especially the structural differences between the hCG β 3-loop and LH β 3-loop that are making the 8G5 antibody able to discriminate between these two different epitopes.

The results presented above indicated the presence of structural differences between the three different peptides of peptides. In order to be able to investigate these peptides via NMR spectroscopy, it was crucial to assign their respective TOCSY spectra. The side-chain assignments using 1D ^1H were fairly straightforward for all three peptides. In the present study it was crucial to study NOE connectivities due to the presence of amino acid redundancy, but also to determine the differences in structure in all three peptides. Unfortunately, this was the hardest task during this project. NOESY experiments would not lead to any sequence-specific NOE at 700 MHz (Figure A.12). This was possibly due to the molecular size of these peptides ($\approx 1,500$ Da) that made them fall in the inconvenient regime where the NOE goes through a null (Figure A.13). Several attempts at getting out of this troublesome regime were considered. Changes in temperatures and magnetic field were investigated. A 600 MHz magnetic field, instead of the usual 700 MHz

field, was tested. It did not lead to any NOEs. As ω_0 was lowered, the tumbling rate must have been reduced too. Since, the NOEs were not visible, the initial tumbling rate at 700 MHz must have been very close to the zero point and dropping the field by 100 MHz must have not had a huge impact on the NOE intensity. Temperature rise (i.e. 37°), or drop (i.e. 15°C), did not resolve the problematic. Hence, ROESY experiment was used to solve this complication.

The preliminary data revealed that MOL179 was highly unstructured, as very little ROEs were observed compared to MOL174 and MOL180 (Figure A.15). If there was a hydrogen clasp between Asp77 and Val77 of hCG β 3 loop, we should have been able to observe NOEs or ROEs between these two residues. Among all three proteins, only MOL174 featured a long-range NOE between amino acids Arg74 and Val79. Between the few observed NOEs, one of them were consistent with the presence of β -turns described by Wuthrich (1986). In an ideal case the segment P-R-G-V, where the β -turn was expected, should manifest strong H_N - H_N NOE between residues R74-V76, and G75-V76, and H_α - H_N NOE between residue R74-G75 (Wuthrich 1986). Only H_α - H_N NOE between R74- H_α and G75- H_N was observed in MOL174, MOL179 and MOL180. Moreover, additional NOEs were only observed for MOL174 and MOL179. These two were also the only peptides showing binding to 8G5 (Gregor et al. 2011) (Figure A.2). In this respect, the NMR data does reflect Gregor et al. (2011)'s postulate that the β 3-loop might have an intrinsic self-contained secondary structure regardless of the rest of the protein, and that its local structure could provide the basis for 8G5 recognition and binding. Deviations of observed chemical shifts from random coil values were indicative of the presence

of secondary structure (Figure A.16). The result suggested the existence of a β -turn like structure in MOL174 and MOL180, and the presence of unordered structure in MOL179 (Figure A.16). The only difference between MOL174 and MOL180 is that the β -turn seems to be defined by different sequences (i.e. MOL174 β -turn seems to span over R-L-P-G residues, while MOL179 β -turn seems to span over P-G-C-P-R-G residues). The different CD data obtained by (Gregor et al. 2011) were indicative of a more disordered structure, thus if turns were present they would exist in an outnumbered group of open conformation (Gregor et al. 2011). In fact this problem was also noticed in the conformational studies of these peptides via NMR. Small numbers of NOEs observed (Figure A.15) and the high $^3J_{\text{H}_\text{N}\text{H}_\alpha}$ -coupling constant values (Table A.2) were observed, and appeared to also corroborate the results obtained with CD. This is due to NMR data reflecting a population-weighted average over all structures present in an ensemble of peptide conformers in solution.



Figure A.18: Data summary. NOEs between two different residues (brackets) and potential β -turn sequences (highlighted in orange) are mapped onto the primary structure of the peptides.

To conclude, this NMR study showed that the interpretation of NMR data to give some insight on peptide conformations was difficult due to flexibility of these polypeptides. However, in view of the above data, the study was able to deliver some light on the structural features associated with β 3-loop epitopic

peptides, which might contribute to 8G5 recognition and binding. Still, we did not observe NOEs between Arg77 and Val7 suggesting the formation of a hydrogen clasp between these two residues, hence the presence of a β -turn like structure in hCG β 3 loop. Despite the fact that we could not prove the direct impact of one-point mutation of residue 77 in 8G5 recognition and binding, the difference in structural features between the hCG and LH β 3-loops are important.

Appendix B

Chemical shift assignment of hCG peptides

B.1 Chemical shift assignment of MOL174 peptide.

All chemical shifts are listed in ppm and referenced to residual water signal. All the chemical shifts that are not present in an amino acid are marked in black and missing assignments with a "-".

Residue Number	MOL174 (N)	NH	δ NH	HA	HB1	HB2	HG1	HG2	γ CH3/2	δ CH3/2	HD1	HD2	ϵ NH	2H	4H
66	Ser	8.333		4.562	3.75	-									
67	Ile	-		4.157	1.777		1.396	1.113	1.079	0.833					
68	Arg	8.377		4.25	1.812	1.718			1.625	3.147			-		
69	Leu	8.347		3.946	1.549		0.944			0.863					
70	Pro			4.341	2.234	1.99			1.883		3.759	3.616			
71	Gly	8.399		3.869											
72	Cys	8.067		4.728	2.838	-									
73	Pro			-	2.222	1.946					3.76	-			
74	Arg	8.37		4.313	1.812	1.718			1.873				8.094		
75	Gly	8.31		3.891					1.514	3.123					
76	Val	7.885		4.063	1.99				0.828						
77	Asn	8.425	6.839/7.53	4.925	2.759	2.605									
78	Pro			4.38	2.213	1.954					3.751	3.672			
79	Val	7.867		4.028	1.956				1.866						
80	Val	7.867		4.028	1.956				0.794						

B.2 Chemical shift assignment of MOL179 peptide.

All chemical shifts are listed in ppm and referenced to residual water signal. All the chemical shifts that are not present in an amino acid are marked in black and missing assignments with a "-".

Residue Number	MOL179 (D)	NH	δ NH	HA	HB1	HB2	HG1	HG2	γ CH3/2	δ CH3/2	HD1	HD2	ϵ NH	2H	4H
86	Ser	8,052		4.729	2.835	-									
87	Ile	8,476		4.175	1.769		1.391	1.113	1.079	0.827					
88	Arg	8,374		4.299	-	1.687			1.544	3.123			7.108		
89	Leu	8,349		4.553	-		1.517			0.864					
90	Pro			4.332	2.23	1.963			-		3.76	3.593			
91	Gly	8,369		3.871											
92	Cys	8,443		4.873	2.801	2.587									
93	Pro			-	2.203	1.935			-		3.672	-			
94	Arg	8,355		4.245	-	1.768			1.663	3.148			7.141		
95	Gly	8,311		3.886											
96	Val	7,91		4.066	2.03				0.821						
97	Asp	8,009		-	2.705	2.546									
98	Pro			4.374	2.209	1.932			-		3.717	3.632			
99	Val	8,105		4.024	1.997				0.872						
100	Val	7,873		4.059	1.99				0.814						

B.3 Chemical shift assignment of MOL180 peptide.

All chemical shifts are listed in ppm and referenced to residual water signal. All the chemical shifts that are not present in an amino acid are marked in black and missing assignments with a "-".

Residue Number	MOL180 (H)	NH	δ NH	HA	HB1	HB2	HG1	HG2	γ CH3/2	δ CH3/2	HD1	HD2	ϵ NH	2H	4H
66	Ser	7.817		4.52	2.765	-									
67	Ile	-		4.144	1.769		1.396	1.113	-	0.833					
68	Arg	8.384		4.249	1.806	1.708			1.594	3.138			7.089		
69	Leu	8.334		4.561	-		1.53			0.854					
70	Pro			4.341	2.226	1.92			1.879		3.766	3.599			
71	Gly	8.41		3.87											
72	Cys	8.091		4.537	2.942	2.822									
73	Pro			4.364	2.228	1.947			-		3.75	3.659			
74	Arg	8.37		4.305	1.716	1.641			1.575	3.117			-		
75	Gly	8.324		3.88											
76	Val	7.844		4.01	1.928				0.772					7.918	6.973
77	His	8.425		4.891	3.108	2.99									
78	Pro			4.396	2.213	1.954					3.72	3.506			
79	Val	8.261		4.064	2.018				1.848						
80	Val	7.615		3.99	1.986				0.896						
									0.818						

Bibliography

- Angulo, J. and Nieto, P. M. (2011) STD-NMR: application to transient interactions between biomolecules a quantitative approach. *European Biophysics Journal* 40.12, pp. 1357–1369.
- Angulo, J., Enríquez-Navas, P. M., and Nieto, P. M. (2010) Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chemistry-A European Journal* 16.26, pp. 7803–7812.
- Ataie, N. J., Hoang, Q. Q., Zahniser, M. P., Tu, Y., Milne, A., Petsko, G. A., and Ringe, D. (2008) Zinc Coordination Geometry and Ligand Binding Affinity: The Structural and Kinetic Analysis of the Second-Shell Serine 228 Residue and the Methionine 180 Residue of the Aminopeptidase from *Vibrio proteolyticus*. *Biochemistry* 47.29, pp. 7673–7683.
- Bagshawe, A. F., Garsia, R. J., Baumgart, K., Astbury, L., et al. (1990) IgM serum antibodies to phenolic glycolipid-I and clinical leprosy: two years' observation in a community with hyperendemic leprosy. *International Journal of Leprosy and Other Mycobacterial Diseases* 58.1, pp. 25–30.

- Bagshawe, K. (1987) Radiochemotherapy with ^{125}I -5-iodo-2-deoxyuridine. *Cancer treatment reviews* 14.3, pp. 397–399.
- Bagshawe, K. (1989) The First Bagshawe lecture. Towards generating cytotoxic agents at cancer sites. *British journal of cancer* 60.3, p. 275.
- Bagshawe, K. D., Sharma, K, Southall, P. J., Boden, J. A., Boxer, G. M., Patridge, T. A., Antoniow, P., and Pedley, R. B. (1991) Selective uptake of toxic nucleoside (^{125}I UdR) by resistant cancer. *The British journal of radiology* 64.757, pp. 37–44.
- Barinka, C., Šácha, P., Sklenář, J., Man, P., Bezouška, K., Slusher, B. S., and Konvalinka, J. (2004) Identification of the N-glycosylation sites on glutamate carboxypeptidase II necessary for proteolytic activity. *Protein science* 13.6, pp. 1627–1635.
- Bax, A. and Davis, D. G. (1985) Practical aspects of two-dimensional transverse NOE spectroscopy. *Journal of Magnetic Resonance (1969)* 63.1, pp. 207–213.
- Bertino, J. and Hillcoat, B. (1968) Regulation of dihydrofolate reductase and other folate-requiring enzymes. *Advances in enzyme regulation* 6, pp. 335–349.
- Bhatia, J, Sharma, S. K., Chester, K. A., Pedley, R. B., Boden, R. W., Read, D. A., Boxer, G. M., Michael, N. P., and Begent, R. H. (2000) Catalytic activity of an in vivo tumor targeted anti-CEA scFv::carboxypeptidase G2 fusion protein. *International Journal of Cancer* 85.4, pp. 571–577. ISSN: 0020-7136 (Print); 0020-7136 (Linking).
- Biagini, A and Puigserver, A (2001) Sequence analysis of the aminoacylase-1 family. A new proposed signature for metalloexopeptidases. *Comparative*

Biochemistry and Physiology Part B: Biochemistry and Molecular Biology

128.3, pp. 469–481.

Bienvenue, D. L., Gilner, D. M., Davis, R. S., Bennett, B., and Holz, R. C.

(2003) Substrate specificity, metal binding properties, and spectroscopic characterization of the DapE-encoded N-succinyl-L, L-diaminopimelic acid desuccinylase from *Haemophilus influenzae*. *Biochemistry* 42.36, pp. 10756–10763.

Bieri, M., Kwan, A. H., Mobli, M., King, G. F., Mackay, J. P., and

Gooley, P. R. (2011) Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. *Febs Journal* 278.5, pp. 704–715.

Birken, S., Armstrongs, E. G., Gawinowicz Kolks, M. A., Cole, L. A., Agosto,

G. M., Krichevsky, A., Vaitukaitis, J. L., and Canfield, R. E. (1988) Structure of the Human Chorionic Gonadotropin β -Subunit Fragment from Pregnancy Urine*. *Endocrinology* 123.1, pp. 572–583.

Bodenhausen, G. and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR

by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* 69.1, pp. 185–189.

Boistelle, R and Astier, J. (1988) Crystallization mechanisms in solution. *Journal*

of Crystal Growth 90.1, pp. 14–30.

Bothner-By, A. A., Stephens, R., Lee, J., Warren, C. D., and Jeanloz, R. (1984)

Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *Journal of the American Chemical Society* 106.3, pp. 811–813.

- Braunschweiler, L. and Ernst, R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *Journal of Magnetic Resonance* 53.3, pp. 521–528.
- Bryant, R. G. (1983) The NMR time scale. *Journal of Chemical Education* 60.11, p. 933.
- Bzymek, K. P. and Holz, R. C. (2004) The catalytic role of glutamate 151 in the leucine aminopeptidase from *Aeromonas proteolytica*. *Journal of Biological Chemistry* 279.30, pp. 31018–31025.
- Cain, J. A., Solis, N., and Cordwell, S. J. (2014) Beyond gene expression: the impact of protein post-translational modifications in bacteria. *Journal of proteomics* 97, pp. 265–286.
- Cala, O., Guillièrre, F., and Krimm, I. (2014) NMR-based analysis of protein-ligand interactions. *Analytical and bioanalytical chemistry* 406.4, pp. 943–956.
- Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., and Skelton, N. J. (1995) *Protein NMR spectroscopy: principles and practice*. Academic Press.
- Chabner, B., Johns, D., and Bertino, J. (1972) Enzymatic cleavage of methotrexate provides a method for prevention of drug toxicity.
- Chen, X., Qin, S., Chen, S., Li, J., Li, L., Wang, Z., Wang, Q., Lin, J., Yang, C., and Shui, W. (2015) A Ligand-observed Mass Spectrometry Approach Integrated into the Fragment Based Lead Discovery Pipeline. *Scientific reports* 5.

- Chester, K. A., Baker, M., and Mayer, A. (2005) Overcoming the immunologic response to foreign enzymes in cancer therapy. *Expert review of clinical immunology* 1.4, pp. 549–559.
- Chevrier, B., Schalk, C., D’Orchymont, H., Rondeau, J.-M., Moras, D., and Tarnus, C. (1994) Crystal structure of *Aeromonas proteolytica* aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. *Structure* 2.4, pp. 283–291.
- Choi, J. and Lee, S. (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Applied microbiology and biotechnology* 64.5, pp. 625–635.
- Ciulli, A., Williams, G., Smith, A. G., Blundell, T. L., and Abell, C. (2006) Probing hot spots at protein-ligand binding sites: a fragment-based approach using biophysical methods. *Journal of medicinal chemistry* 49.16, pp. 4992–5000.
- Claridge, T. D. (2008) *High-resolution NMR techniques in organic chemistry*. Vol. 27. Newnes.
- Clarkson, J and Campbell, I. (2003) Studies of protein-ligand interactions by NMR. *Biochemical Society Transactions* 31.5, pp. 1006–1009.
- Consalvi, V., Chiaraluce, R., Giangiacomo, L., Scandurra, R., Christova, P., Karshikoff, A., Knapp, S., and Ladenstein, R. (2000) Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile *Thermotoga maritima*. *Protein engineering* 13.7, pp. 501–507.

- Crider, K. S., Yang, T. P., Berry, R. J., and Bailey, L. B. (2012) Folate and DNA methylation: a review of molecular mechanisms and the evidence for folates role. *Advances in Nutrition: An International Review Journal* 3.1, pp. 21–38.
- Cutting, B., Shelke, S. V., Dragic, Z., Wagner, B., Gathje, H., Kelm, S., and Ernst, B. (2007) Sensitivity enhancement in saturation transfer difference (STD) experiments through optimized excitation schemes. *Magnetic Resonance in Chemistry* 45.9, pp. 720–724.
- Czisch, M and Boelens, R (1998) Sensitivity enhancement in the TROSY experiment. *Journal of Magnetic Resonance* 134.1, pp. 158–160.
- Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., and Stockman, B. (2001) WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *Journal of biomolecular NMR* 21.4, pp. 349–359.
- Darwish, I. A. (2006) Immunoassay methods and their applications in pharmaceutical analysis: basic methodology and recent advances. *International Journal of Biomedical Science* 2, pp. 217–235.
- DeAngelis, L. M., Tong, W. P., Lin, S., Fleisher, M., and Bertino, J. R. (1996) Carboxypeptidase G2 rescue after high-dose methotrexate. *Journal of clinical oncology* 14.7, pp. 2145–2149.
- Demarco, A., Llinás, M., and Wüthrich, K. (1978) Analysis of the ^1H -NMR spectra of ferrichrome peptides. I. The non-amide protons. *Biopolymers* 17.3, pp. 617–636.
- Derome, A. E. (2013) *Modern NMR techniques for chemistry research*. Elsevier.

- Dötsch, V. and Withers, R. S. (2005) For use in spectrometer, the buffer solution comprising a primary buffer and a titrating agent with low ionic mobility, without reducing the salt content, increase sensitivity. US Patent 6,958,244.
- Ducat, T., Declerck, N., Gostan, T., Kochoyan, M., and Demene, H. (2006) Rapid determination of protein solubility and stability conditions for NMR studies using incomplete factorial design. *Journal of biomolecular NMR* 34.3, pp. 137–151.
- Ducruix, A and Giege, R (1999) An introduction to the crystallogenesi s of biological macromolecules. *Crystallization of Nucleic Acids and Proteins. A Practical Approach*, pp. 1–16.
- Elliott, M. M., Kardana, A., Lustbader, J. W., and Cole, L. A. (1997) Carbohydrate and peptide structure of the α - and β -subunits of human chorionic gonadotropin from normal and aberrant pregnancy and choriocarcinoma. *Endocrine* 7.1, pp. 15–32.
- Fernández, C. and Wider, G. (2003) TROSY in NMR studies of the structure and function of large biological macromolecules. *Current opinion in structural biology* 13.5, pp. 570–580.
- Fielding, L. (2007) NMR methods for the determination of protein–ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy* 51.4, pp. 219–242.
- Fielding, L., Rutherford, S., and Fletcher, D. (2005) Determination of protein–ligand binding affinity by NMR: observations from serum albumin model systems. *Magnetic Resonance in Chemistry* 43.6, pp. 463–470.

- Goda, S. K., Rashidi, F. A. B., Fakhro, A. A., and Al-obaidli, A. (2009) Functional Overexpression and Purification of a Codon Optimized Synthetic Glucarpidase (Carboxypeptidase G2) in *Escherichia coli*. *The protein journal* 28.9-10, pp. 435–442.
- Goldflam, M., Tarragó, T., Gairí, M., and Giralt, E. (2012) “NMR studies of protein-ligand interactions”. *Protein NMR techniques*. Springer, pp. 233–259.
- Goto, N. K. and Kay, L. E. (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Current opinion in structural biology* 10.5, pp. 585–592.
- Greenfield, N. J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature protocols* 1.6, pp. 2527–2535.
- Gregor, C. R., Cerasoli, E., Schouten, J., Ravi, J., Slootstra, J., Horgan, A., Martyna, G. J., Ryadnov, M. G., Davis, P., and Crain, J. (2011) Antibody Recognition of a Human Chorionic Gonadotropin Epitope (hCG β 66-80) Depends on Local Structure Retained in the Free Peptide. *Journal of Biological Chemistry* 286.28, pp. 25016–25026.
- Griffey, R. T., Trent, C. J., Bavolek, R. A., Keeperman, J. B., Sampson, C., and Poirier, R. F. (2013) Hook-like effect causes false-negative point-of-care urine pregnancy testing in emergency patients. *The Journal of emergency medicine* 44.1, pp. 155–160.
- Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology* 166.4, pp. 557–580.

- Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S.-J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., et al. (2007) Crystallization of protein–ligand complexes. *Acta Crystallographica Section D: Biological Crystallography* 63.1, pp. 72–79.
- Heller, M. and Kessler, H. (2001) NMR spectroscopy in drug design. *Pure and Applied Chemistry* 73.9, pp. 1429–1436.
- Hwang, T.-L. and Shaka, A. (1995) Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A* 112.2, pp. 275–279.
- Johnsson, B., Löfås, S., and Lindquist, G. (1991) Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Analytical biochemistry* 198.2, pp. 268–277.
- Jozic, D., Bourenkow, G., Bartunik, H., Scholze, H., Dive, V., Henrich, B., Huber, R., Bode, W., and Maskos, K. (2002) Crystal structure of the dinuclear zinc aminopeptidase PepV from *Lactobacillus delbrueckii* unravels its preference for dipeptides. *Structure* 10.8, pp. 1097–1106.
- Kalk, A and Berendsen, H. (1976) Proton magnetic relaxation and spin diffusion in proteins. *Journal of Magnetic Resonance (1969)* 24.3, pp. 343–366.
- Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., and Waugh, D. S. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein engineering* 14.12, pp. 993–1000.

- Kay, L. E. (2005) NMR studies of protein structure and dynamics. *Journal of Magnetic Resonance* 173.2, pp. 193–207.
- Keeler, J. (2011) *Understanding NMR spectroscopy*. John Wiley & Sons.
- Kelly, A. E., Ou, H. D., Withers, R., and Dötsch, V. (2002) Low-conductivity buffers for high-sensitivity NMR measurements. *Journal of the American Chemical Society* 124.40, pp. 12013–12019.
- Khan, T. H., Eno-Amooquaye, E. A., Searle, F., Browne, P. J., Osborn, H. M., and Burke, P. J. (1999) Novel inhibitors of carboxypeptidase G2 (CPG2): potential use in antibody-directed enzyme prodrug therapy. *Journal of medicinal chemistry* 42.6, pp. 951–956.
- Kieffer, B., Homans, S., and Jahnke, W (2011) Nuclear magnetic resonance of ligand binding to proteins. *Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection, Measurement and Modelling* 22, p. 15.
- Klusák, V., Barinka, C., Plechanovová, A., Mlcochová, P., Konvalinka, J., Rulíšek, L. r. Rulíšek, and Lubkowski, J. (2009) Reaction mechanism of glutamate carboxypeptidase II revealed by mutagenesis, X-ray crystallography, and computational methods. *Biochemistry* 48.19, pp. 4126–4138.
- Kneller, D. and Kuntz, I. (1993) “UCSF Sparky-an NMR display, annotation and assignment tool”. *Journal of Cellular Biochemistry*. Ed. by I. Wiley-Liss, pp. 254–254.
- Kumar, A., Ernst, R., and Wüthrich, K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological

- macromolecules. *Biochemical and biophysical research communications* 95.1, pp. 1–6.
- Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227.5259, pp. 680–685.
- Lapthorn, A., Harris, D., Littlejohn, A., Lustbader, J., Canfield, R., Machin, K., Morgan, F., and Isaacs, N. (1994) Crystal structure of human chorionic gonadotropin. *Nature* 369.6480, pp. 455–461.
- Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31.8, pp. 1325–1327.
- Levitt, M. H. (2001) *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons.
- Levy, C. C. and Goldman, P. (1967) The enzymatic hydrolysis of methotrexate and folic acid. *Journal of Biological Chemistry* 242.12, pp. 2933–2938.
- Lindner, H. A., Lunin, V. V., Alary, A., Hecker, R., Cygler, M., and Ménard, R. (2003) Essential roles of zinc ligation and enzyme dimerization for catalysis in the aminoacylase-1/M20 family. *Journal of Biological Chemistry* 278.45, pp. 44496–44504.
- Lowther, W. T. and Matthews, B. W. (2002) Metalloaminopeptidases: common functional themes in disparate structural surroundings. *Chemical Reviews* 102.12, pp. 4581–4608.
- Ludwig, C. and Guenther, U. L. (2009) Ligand based NMR methods for drug discovery. *Frontiers in Bioscience* 14.4565-74, p. 24.

- Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *Journal of biomolecular NMR* 20.1, pp. 71–75.
- Matthews, S. (2004) “Perdeuteration/Site-Specific Protonation Approaches for High-Molecular-Weight Proteins”. *Protein NMR Techniques*. Springer, pp. 35–45.
- Meissner, A., SCHULTE-HERBRÜGGEN, T., Briand, J., and SØRENSEN, O. W. (1998) Double spin-state-selective coherence transfer. Application for two-dimensional selection of multiplet components with long transverse relaxation times. *Molecular Physics* 95.6, pp. 1137–1142.
- Minton, N. P., Atkinson, T., and Sherwood, R. F. (1983) Molecular cloning of the *Pseudomonas* carboxypeptidase G2 gene and its expression in *Escherichia coli* and *Pseudomonas putida*. *Journal of bacteriology* 156.3, pp. 1222–1227.
- Minton, N. P., Atkinson, T., Bruton, C. J., and Sherwood, R. F. (1984) The complete nucleotide sequence of the *Pseudomonas* gene coding for carboxypeptidase G2. *Gene* 31.1, pp. 31–38.
- Miroux, B. and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology* 260.3, pp. 289–298.
- Murray, V., Huang, Y., Chen, J., Wang, J., and Li, Q. (2012) “A novel bacterial expression method with optimized parameters for very high yield production of triple-labeled proteins”. *Protein NMR Techniques*. Springer, pp. 1–18.
- Nietlispach, D., Clowes, R. T., Broadhurst, R. W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P. J., and Laue, E. D. (1996) An

- approach to the structure determination of larger proteins using triple resonance NMR experiments in conjunction with random fractional deuteration. *Journal of the American Chemical Society* 118.2, pp. 407–415.
- O'Connor, J. F., Birken, S., Lustbader, J. W., Krichevsky, A., Chen, Y., and Canfield, R. E. (1994) Recent Advances in the Chemistry and Immunochemistry of Human Chorionic Gonadotropin: Impact on Clinical Measurements*. *Endocrine reviews* 15.5, pp. 650–683.
- Ollerenshaw, J. E., Tugarinov, V., and Kay, L. E. (2003) Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* 41.10, pp. 843–852.
- Osborn, M. and Huennekens, F. (1958) Enzymatic reduction of dihydrofolic acid. *Journal of Biological Chemistry* 233.4, pp. 969–974.
- Pace, C, Grimsley, G. R., Thomas, S. T., and Makhatadze, G. I. (1999) Heat capacity change for ribonuclease A folding. *Protein science* 8.07, pp. 1500–1504.
- Pardi, A, Billeter, M, and Wüthrich, K (1984) Calibration of the angular dependence of the amide proton- $C\ \alpha$ proton coupling constants, $3\ J\ H_N\alpha$, in a globular protein: Use of $3JH_N\alpha$ for identification of helical secondary structure. *Journal of molecular biology* 180.3, pp. 741–751.
- Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997) Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* 94.23, pp. 12366–12371.

- Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C -labeled proteins. *Journal of the American Chemical Society* 120.25, pp. 6394–6400.
- Pervushin, K., Braun, D., Fernández, C., and Wüthrich, K. (2000) ^{15}N , ^1H / ^{13}C , ^1H -TROSY for simultaneous detection of backbone ^{15}N - ^1H , aromatic ^{13}C - ^1H and side-chain ^{15}N - $^1\text{H}_2$ correlations in large proteins. *Journal of biomolecular NMR* 17.3, pp. 195–202.
- Piantini, U, Sorensen, O., and Ernst, R. R. (1982) Multiple quantum filters for elucidating NMR coupling networks. *Journal of the American Chemical Society* 104.24, pp. 6800–6801.
- Pierce, J. G. (1971) Eli Lilly Lecture: The Subunits of Pituitary Thyrotropin Their Relationship to Other Glycoprotein Hormones 1. *Endocrinology* 89.6, pp. 1331–1344.
- Piotto, M., Saudek, V., and Sklenář, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR* 2.6, pp. 661–665.
- Privalov, P. L. (1997) Thermodynamics of protein folding. *The Journal of Chemical Thermodynamics* 29.4, pp. 447–474.
- Privalov, P. and Khechinashvili, N. (1974) A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *Journal of molecular biology* 86.3, pp. 665–684.

- Rance, M. (1987) Improved techniques for homonuclear rotating-frame and isotropic mixing experiments. *Journal of Magnetic Resonance* 74.3, pp. 557–564.
- Rance, M., Loria, J. P., and Palmer, A. G. (1999) Sensitivity improvement of transverse relaxation-optimized spectroscopy. *Journal of Magnetic Resonance* 136.1, pp. 92–101.
- Rance, M., Sørensen, O., Bodenhausen, G., Wagner, G., Ernst, R., and Wüthrich, K (1983) Improved spectral resolution in COSY 1 H NMR spectra of proteins via double quantum filtering. *Biochemical and biophysical research communications* 117.2, pp. 479–485.
- Remaut, H., Bompard-Gilles, C., Goffin, C., Frère, J.-M., and Van Beeumen, J. (2001) Structure of the *Bacillus subtilis* D-aminopeptidase DppA reveals a novel self-compartmentalizing protease. *Nature Structural & Molecular Biology* 8.8, pp. 674–678.
- Richardson, J. S. (1981) *The anatomy and taxonomy of protein structure*. Vol. 34. Academic Press.
- Roberts, G. and Lian, L.-Y. (2011) *Protein NMR spectroscopy: practical techniques and applications*. John Wiley & Sons.
- Rowell, S., Paupit, R. A., Tucker, A. D., Melton, R. G., Blow, D. M., and Brick, P. (1997) Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure* 5.3, pp. 337–347.
- Sanbrook, J, Fritsch, E., and Maniatis, T (1989) Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY*.

- Schürer, G., Lanig, H., and Clark, T. (2004) *Aeromonas proteolytica* aminopeptidase: an investigation of the mode of action using a quantum mechanical/molecular mechanical approach. *Biochemistry* 43.18, pp. 5414–5427.
- Schwartz, S., Borner, K., Müller, K., Martus, P., Fischer, L., Korfel, A., Auton, T., and Thiel, E. (2007) Glucarpidase (carboxypeptidase g2) intervention in adult and elderly cancer patients with renal dysfunction and delayed methotrexate elimination after high-dose methotrexate therapy. *The Oncologist* 12.11, pp. 1299–1308.
- Searle, F., Bagshawe, K., Boden, J., Bier, C., Green, A., Pedley, R., Melton, R., and Sherwood, R. (1986) Carboxypeptidase G2 conjugates with localizing antitumor antibodies-potential therapeutic agents. *Tumor Biology* 7.4, pp. 320–320.
- Shaka, A., Keeler, J., and Freeman, R. (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. *Journal of Magnetic Resonance* 53.2, pp. 313–340.
- Shaka, A., Lee, C., and Pines, A. (1988) Iterative schemes for bilinear operators; application to spin decoupling. *Journal of Magnetic Resonance* 77.2, pp. 274–293.
- Sharma, S. K., Bagshawe, K. D., Melton, R. G., and Sherwood, R. F. (1992) Human immune response to monoclonal antibody-enzyme conjugates in ADEPT pilot clinical trial. *Cell biophysics* 21.1-3, pp. 109–120.

- Shaw, M. K. and Ingraham, J. L. (1967) Synthesis of macromolecules by *Escherichia coli* near the minimal temperature for growth. *Journal of bacteriology* 94.1, pp. 157–164.
- Sherwood, R. F., Melton, R. G., Alwan, S. M., and Hughes, P. (1985) Purification and properties of carboxypeptidase G2 from *Pseudomonas* sp. strain RS-16. *European Journal of Biochemistry* 148.3, pp. 447–453.
- Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274.5292, pp. 1531–1534.
- Singh, S. M. and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins. *Journal of bioscience and bioengineering* 99.4, pp. 303–310.
- Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q. (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science* 18.5, pp. 936–948.
- Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) Gradient-tailored water suppression for ^1H - ^{15}N HSQC experiments optimized to retain full sensitivity. *Journal of Magnetic Resonance, Series A* 102.2, pp. 241–245.
- Sørensen, H. P. and Mortensen, K. K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial cell factories* 4.1, p. 1.
- Spencer, D. I., Robson, L., Purdy, D., Whitelegg, N. R., Michael, N. P., Bhatia, J., Sharma, S. K., Rees, A. R., Minton, N. P., Begent, R. H., et al. (2002) A

strategy for mapping and neutralizing conformational immunogenic sites on protein therapeutics. *Proteomics* 2.3, pp. 271–279.

Stark, J. L. and Powers, R. (2012) “Application of NMR and molecular docking in structure-based drug discovery”. *NMR of Proteins and Small Biomolecules*. Springer, pp. 1–34.

Studier, F. W. and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* 189.1, pp. 113–130.

Sudmeier, J., Evelhoch, J., and Jonsson, N.-H. (1980) Dependence of NMR lineshape analysis upon chemical rates and mechanisms: Implications for enzyme histidine titrations. *Journal of Magnetic Resonance* 40.2, pp. 377–390.

Torres, F. E., Recht, M. I., Coyle, J. E., Bruce, R. H., and Williams, G. (2010) Higher throughput calorimetry: opportunities, approaches and challenges. *Current opinion in structural biology* 20.5, pp. 598–605.

Tucker, A., Rowsell, S, Melton, R., and Pauptit, R. (1996) A new crystal form of carboxypeptidase G2 from *Pseudomonas* sp. strain RS-16 which is more amenable to structure determination. *Acta Crystallographica Section D: Biological Crystallography* 52.4, pp. 890–892.

Tugarinov, V., Choy, W.-Y., Orekhov, V. Y., and Kay, L. E. (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proceedings of the National Academy of Sciences of the United States of America* 102.3, pp. 622–627.

- Turra, K. M., Pasqualoto, K. F. M., Ferreira, E. I., and Rando, D. G. (2012) Molecular modeling approach to predict a binding mode for the complex methotrexate-carboxypeptidase G2. *Journal of molecular modeling* 18.5, pp. 1867–1875.
- Unger, T (1999) Crystallization methods. *International University Line, La Jolla, Calif.*
- Valerino, D., Johns, D., Zaharko, D., and Oliverio, V. (1972) Studies of the metabolism of methotrexate by intestinal floraI: Identification and study of biological properties of the metabolite 4-amino-4-deoxy-N10-methylpteroic acid. *Biochemical pharmacology* 21.6, pp. 821–831.
- Vermeer, A. W. and Norde, W. (2000) The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein. *Biophysical Journal* 78.1, pp. 394–404.
- Viegas, A., Manso, J., Nobrega, F. L., and Cabrita, E. J. (2011) Saturation-transfer difference (STD) NMR: a simple and fast method for ligand screening and characterization of protein binding. *Journal of Chemical Education* 88.7, pp. 990–994.
- Waldner, J. C., Lahr, S. J., Edgell, M. H., and Pielak, G. J. (1999) Nonideality and protein thermal denaturation. *Biopolymers* 49.6, pp. 471–479.
- Weickert, M. J., Doherty, D. H., Best, E. A., and Olins, P. O. (1996) Optimization of heterologous protein production in *Escherichia coli*. *Current opinion in biotechnology* 7.5, pp. 494–499.

- Weigelt, J. (1998) Single scan, sensitivity-and gradient-enhanced TROSY for multidimensional NMR experiments. *Journal of the American Chemical Society* 120.41, pp. 10778–10779.
- West, R., Buffy, J. J., Haaf, M., Müller, T., Gehrhus, B., Lappert, M. F., and Apeloig, Y. (1998) Chemical shift tensors and NICS calculations for stable silylenes. *Journal of the American Chemical Society* 120.7, pp. 1639–1640.
- Whitmore, L. and Wallace, B. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic acids research* 32.suppl 2, W668–W673.
- Whitmore, L. and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 89.5, pp. 392–400.
- Williamson, M. P. (2009) Applications of the NOE in Molecular Biology. *Annual reports on NMR spectroscopy* 65, pp. 77–109.
- Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* 73, pp. 1–16.
- Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *Journal of molecular biology* 222.2, pp. 311–333.
- Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of biomolecular NMR* 5.1, pp. 67–81.

- Wouters, M. A. and Husain, A. (2001) Changes in zinc ligation promote remodeling of the active site in the zinc hydrolase superfamily. *Journal of molecular biology* 314.5, pp. 1191–1207.
- Wuthrich, K. (1986) *NMR of proteins and nucleic acids*. Wiley.
- Wuthrich K.thrich, K. (1998) The second decade-into the third millenium. *Nature structural biology* 5.7, pp. 492–495.
- Zhu, G., Kong, X. M., and Sze, K. H. (1999) Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *Journal of biomolecular NMR* 13.1, pp. 77–81.